



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 166995

TO: Manjunath N Rao  
Location: rem-2a01/2c70  
Art Unit: 1652  
Friday, September 23, 2005  
Case Serial Number: 10/777828

From: Toby Port  
Location: Biotech-Chem Library  
REM-1A59  
Phone: 571-272-2523  
[toby.port@uspto.gov](mailto:toby.port@uspto.gov)

### Search Notes

Examiner Rao,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Toby Port  
X22523

This Page Blank (uspto)

STIC-Biotech/ChemLib

---

165996

From: Rao, Manjunath N.  
Sent: Friday, September 16, 2005 2:37 PM  
To: STIC-Biotech/ChemLib  
Subject: 10777828

From: Manjunath N. Rao  
Art Unit 1652, Room 2A01  
Mail Box in Room 2C70  
Phone: 272-0939

Date: 9-16-05

Please search the following as soon as possible for application with serial number

**10/777828**

1. SEQ ID NO: 8 AND Nucleotides 1-115 and nucleotides 428-1011 of SEQ ID NO:8 against all commercial nucleic acid sequence databases, issued patents/published applications nucleic acid sequence database and pending application nucleic acid sequence database. Please provide a print of all results

If you have any questions please call me at the above phone number.

Thanks

9/16/2005

Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2005, 07:38:28 ; Search time 4623.4 Seconds  
(without alignments)  
10595.706 Million cell updates/sec

Title: US-10-777-828-8

Perfect score: 1011

Sequence: 1 ccacctgcgcctctagcat.....aagattgctgcgtgtctga 1011

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hgt.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_to.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1011	100.0	1011	6	B228354	BD228354 UDP-galac
2	1006.2	99.5	10562	9	E38420	E38420 Novel polyp
3	1006.2	99.5	170121	9	AF064860	AF064860 Homo sapi
4	1006.2	99.5	340000	9	HS21C080	AL163280 Homo sapi
5	985.4	97.5	192219	9	RF43002119	AL954227 Pan trogl
6	978.4	96.8	1576	9	AB041416	AB041416 Homo sapi
7	957	94.7	1570	9	AB041415	AB041415 Pan panis
8	955.6	94.5	1565	9	AB041413	AB041413 Homo sapi
9	948.4	93.8	1360	9	AB041412	AB041412 Gorilla g
10	945.6	93.5	1566	9	AB041414	AB041414 Pan trogl
11	930.2	92.0	2494	6	CQ731786	CQ731786 Sequence
12	928.8	91.9	933	9	HS46078	AJ006078 Homo sapi
13	928.6	91.8	2762	9	AB020337	AB020337 Homo sapi
14	928.6	91.8	2775	6	E38419	E38419 Novel polyp
15	928.6	91.8	2921	9	AY372061	AY372061 Homo sapi
16	928.2	91.8	933	9	AF145784	AF145784 Homo sapi
17	926.4	91.6	1579	9	AB041417	AB041417 Pongo pyg
18	771.8	76.3	937	9	AY231145	AY231145 Macaca mu
19	586	58.0	170108	2	AC150794	AC150794 Bos tauru

	20	532.4	52.7	196900	2	AC020851	AC020851 Mus muscu
	21	527.6	52.2	149964	2	AC120145	AC120145 Mus muscu
c	22	527.6	52.2	186956	2	AC120346	AC120346 Mus muscu
	23	525.6	52.0	4933	10	BC057887	BC057887 Mus muscu
	24	525.6	52.0	5069	10	BC051669	BC051669 Mus muscu
	25	522.4	51.7	927	10	AF254738	AF254738 Mus muscu
c	26	323.8	32.0	65400	2	AC109264	AC109264 Mus muscu
	27	269	26.6	97702	2	AC151319	AC151319 Xenopus t
c	28	209.4	20.7	226720	2	BC936311	BC936311 Danio rer
	29	206.2	20.4	2317	5	BC066477	BC066477 Danio rer
	30	185.2	18.3	917	10	AB039137	AB039137 Mus muscu
	31	185.2	18.3	917	10	AB039140	AB039140 Mus muscu
	32	185.2	18.3	917	10	AB039141	AB039141 Mus muscu
	33	185.2	18.3	917	10	AB039142	AB039142 Mus muscu
	34	183.6	18.2	917	10	AB039136	AB039136 Mus muscu
	35	183.6	18.2	917	10	AB039138	AB039138 Mus muscu
	36	183.6	18.2	917	10	AB039143	AB039143 Mus spici
	37	182	18.0	917	10	AB039134	AB039134 Mus muscu
	38	182	18.0	917	10	AB039135	AB039135 Mus muscu
	39	182	18.0	917	10	AB039139	AB039139 Mus muscu
	40	182	18.0	1175	10	AF029790	AF029790 Mus muscu
	41	182	18.0	188273	10	BC936296	BC936296 Mouse DNA
c	42	182	18.0	219368	10	AC098721	AC098721 Mus muscu
c	43	181.2	17.9	207679	2	EX511272	EX511272 Danio rer
c	44	181	17.9	110000	2	AC096079	Continuation (2 of
c	45	181	17.9	231235	2	AC136678	AC136678 Rattus no

ALIGNMENTS

RESULT 1	BD228354	BD228354	1011 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	BD228354	UDP-galactose: beta-N-acetyl-glucosamine beta 1,3				
DEFINITION	BD228354	galactosyltransferases, beta 3 Gal-T5.				
ACCESSION	BD228354.1	GI:33038124				
VERSION	JP 2002530071-A/8.					
KEYWORDS	Homo sapiens (human)					
SOURCE	Homo sapiens					
ORGANISM	Homo sapiens					
REFERENCE	1 (bases 1 to 1011)					
AUTHORS	Clausen,H. and Amado,M.					
TITLE	UDP-galactose: beta-N-acetyl-glucosamine beta 1,3					
JOURNAL	galactosyltransferases, beta 3 Gal-T5					
COMMENT	Patent: JP 2002530071-A 8 17-SEP-2002; HENRIK CLAUSEN OS Homo sapiens (human) PN JP 2002530071-A/8 PD 17-SEP-2002 PF 11-NOV-1999 JP 2000582542 PR 13-NOV-1998 DK PA 199801483 PI HENRIK CLAUSEN,MARGARIDA AMADO PC C12N15/09,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12N9/10,C12N15/10 PC 00,C12N5/00 CC UDP-galactose: beta-N-acetyl-glucosamine beta 1,3 CC galactosyltransferases, beta 3 Gal-T5 CC beta 3 Gal-T5 FH Key Location/Qualifiers FT CDS Location/Qualifiers 1..1011 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"					

ORIGIN

Query Match 100.0%; Score 1011; DB 6; Length 1011;  
Best Local Similarity 100.0%; Pred. No. 7.8e-270;  
Matches 1011; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCTCAGCCTCTAGCATAAACTAGACATCCTCATGCTTTTGGGCTCTAATCATT 60  
Db 1 CCACCTCAGCCTCTAGCATAAACTAGACATCCTCATGCTTTTGGGCTCTAATCATT 60  
QY 61 GGATTTTGTTCCTTTTCAGATGCTTTCCGGAAGATGAGATTGATATATCTGCTTCTG 120  
Db 61 GGATTTTGTTCCTTTTCAGATGCTTTCCGGAAGATGAGATTGATATATCTGCTTCTG 120  
QY 121 GTTCTGGGGGCTCTTTGTTGTTATTTTATGATGTCATCTAAATCTCTTCAAGAACAG 180  
Db 121 GTTCTGGGGGCTCTTTGTTGTTATTTTATGATGTCATCTAAATCTCTTCAAGAACAG 180  
QY 181 TCCTTTGTTTACAAGAACAGAGGGAACCTCTTAAGCTCCAGATACAGACTGCAGGCGAG 240  
Db 181 TCCTTTGTTTACAAGAACAGAGGGAACCTCTTAAGCTCCAGATACAGACTGCAGGCGAG 240  
QY 241 ACACCTCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
Db 241 ACACCTCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
QY 301 GCATCCGGCAGACGTGGGGGAAAGAGAGGAGCGGTGAAGGGAAGCAGCTGAAGACATTC 360  
Db 301 GCATCCGGCAGACGTGGGGGAAAGAGAGGAGCGGTGAAGGGAAGCAGCTGAAGACATTC 360  
QY 361 TTCTCTCTGGGGACCAACAGCAGTGCAGCGGAACAAAGAGGTTGGACAGGAGGCCAG 420  
Db 361 TTCTCTCTGGGGACCAACAGCAGTGCAGCGGAACAAAGAGGTTGGACAGGAGGCCAG 420  
QY 421 CGACACGGGACATTTATCAGAAGGATTTCTTAGAGCTCTATTACAATCTGACCCCTGAAG 480  
Db 421 CGACACGGGACATTTATCAGAAGGATTTCTTAGAGCTCTATTACAATCTGACCCCTGAAG 480  
QY 481 ACCATGATGGGCAATAGATGGTTCATCGCTTTTGTCTCAGCGGCGCTTTGTGATGAAA 540  
Db 481 ACCATGATGGGCAATAGATGGTTCATCGCTTTTGTCTCAGCGGCGCTTTGTGATGAAA 540  
QY 541 ACAGACTCAGACATGTTTCAATCAATGTTGACTATCTGACTGACTGCTTCTGAAGAAAAC 600  
Db 541 ACAGACTCAGACATGTTTCAATCAATGTTGACTATCTGACTGACTGCTTCTGAAGAAAAC 600  
QY 601 AGAACCAACAGGTTTTTCACTGGCTTTCTTGAACACTCAATGATTTTCCCATCAGGCGCA 660  
Db 601 AGAACCAACAGGTTTTTCACTGGCTTTCTTGAACACTCAATGATTTTCCCATCAGGCGCA 660  
QY 661 TTCAGCAAGTGGTTTGTTCAGTAAATCTGAATATCCGTGGGACAGTACCCACCATTTGCG 720  
Db 661 TTCAGCAAGTGGTTTGTTCAGTAAATCTGAATATCCGTGGGACAGTACCCACCATTTGCG 720  
QY 721 TCCGGCACCGGCTACGTGTTTCTGGGACGTTGGGAGTCAAGTGTACAATGTCTCCAAAG 780  
Db 721 TCCGGCACCGGCTACGTGTTTCTGGGACGTTGGGAGTCAAGTGTACAATGTCTCCAAAG 780  
QY 781 AGCGTCCCATACATTTAAACTGGAACAGCTGTTTGTGGGGCTCTGCTCGAAGGCTGAAC 840  
Db 781 AGCGTCCCATACATTTAAACTGGAACAGCTGTTTGTGGGGCTCTGCTCGAAGGCTGAAC 840  
QY 841 ATCAGATGGAGGAGCTCCATCCAGCGGACCTTTTTCAGGGGGGCTTACGCTTCTCC 900  
Db 841 ATCAGATGGAGGAGCTCCATCCAGCGGACCTTTTTCAGGGGGGCTTACGCTTCTCC 900  
QY 901 GTATGCTCTTCAGAGGATCGTGGGCTGCCACTTTCATCAAGCCCTCGGACTCTCTGGAC 960  
Db 901 GTATGCTCTTCAGAGGATCGTGGGCTGCCACTTTCATCAAGCCCTCGGACTCTCTGGAC 960  
QY 961 TACTGGCAGGCTCTAGAGAAATTCGCGGGGGGAAGATTGTTCGCTCTGTCTGA 1011  
Db 961 TACTGGCAGGCTCTAGAGAAATTCGCGGGGGGAAGATTGTTCGCTCTGTCTGA 1011

RESULT 2  
E38420  
LOCUS  
DEFINITION Novel polypeptide.  
10562 bp DNA linear PAT 31-JAN-2002

ACCSSION E38420  
VERSION E38420.1 GI:18626994  
KEYWORDS JP 2000245464-A/2.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mammalia; Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 10562)  
AUTHORS Narimatsu, H., Iishiki, S., Togayauchi, A. and Sasaki, K.  
TITLE Novel polypeptide  
JOURNAL Patent: JP 2000245464-A 2 12-SEP-2000;  
KYOWA HAKKO KOGYO CO LTD  
COMMENT OS Homo sapiens (human)  
PN JP 2000245464-A/2  
PD 12-SEP-2000  
PF 25-FEB-1999 JP 1999047571  
PR HISASHI NARIMATSU, SOICHIRO ISSHIKI, AKIRA TOGAYAUCHI, PI  
KATSUTOSHI SASAKI  
PC C12N15/09, A01K67/027, C12N1/21, C12N5/10, C12N9/10, C12P19/00, PC  
C12P21/02,  
PC C12P21/08, C12Q1/68, G01N33/53// (C12N1/21, C12R1:185), (C12N5/10,  
C12R1:91),  
PC (C12P21/02, C12R1:185), (C12P21/02, C12R1:91), C12N15/00, C12N5/00,  
PC (C12N5/00, C12R1:91)  
CC  
FH Key Location/Qualifiers  
FT promoter (1)..(5000)  
FT exon (5001)..(5140)  
FT exon (5001)..(5273)  
FT exon (5459)..(5567)  
FT exon (7427)..(7586)  
FT exon (8234)..(10562).  
FEATURES  
source  
1..10562  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 99.5%; Score 1006.2; DB 6; Length 10562;  
Best Local Similarity 99.7%; Pred. No. 1.9e-268;  
Matches 1008; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CCACCTCAGCCTCTAGCATAAACTAGACATCCTCATGCTTTTGGGCTCTAATCATT 60  
Db 8156 CCACCTCAGCCTCTAGCATAAACTAGACATCCTCATGCTTTTGGGCTCTAATCATT 8215  
QY 61 GGATTTTGTTCCTTTTCAGATGCTTTCCGGAAGATGAGATTGATATATCTGCTTCTG 120  
Db 8216 GGATTTTGTTCCTTTTCAGATGCTTTCCGGAAGATGAGATTGATATATCTGCTTCTG 8275  
QY 121 GTTCTGGGGGCTCTTTGTTGTTATTTTATGATGTCATCTAAATCTCTTCAAGAACAG 180  
Db 8276 GTTCTGGGGGCTCTTTGTTGTTATTTTATGATGTCATCTAAATCTCTTCAAGAACAG 8335  
QY 181 TCCTTTGTTTACAAGAACAGCGGAACTTCTTAAGCTCCAGATACAGACTGCAGGCGAG 240  
Db 8336 TCCTTTGTTTACAAGAACAGCGGAACTTCTTAAGCTCCAGATACAGACTGCAGGCGAG 8395  
QY 241 ACACCTCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
Db 8396 ACACCTCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8455  
QY 301 GCATCCGGCAGACGTGGGGGAAAGAGAGGAGCGGTGAAGGGAAGCAGCTGAAGACATTC 360  
Db 8456 GCATCCGGCAGACGTGGGGGAAAGAGAGGAGCGGTGAAGGGAAGCAGCTGAAGACATTC 8515  
QY 361 TTCTCTCTGGGGACCAACAGCAGTGCAGCGGAACAAAGAGGTTGGACAGGAGGCCAG 420  
Db 8516 TTCTCTCTGGGGACCAACAGCAGTGCAGCGGAACAAAGAGGTTGGACAGGAGGCCAG 8575  
QY 421 CGACACGGGACATTTATCAGAAGGATTTCTTAGAGCTCTATTACAATCTGACCCCTGAAG 480

```
Db 8576 CGACACGGGACATTTATCCAGAGGATTTCTAGACGCTATTATCAATCTGACCCCTGAAG 8635
Qy 481 ACCATGATGGGCATAGATGGGTCCATCGCTTTCTCTCAGGCGGGCTTTGTGATGAAA 540
Db 8636 ACCATGATGGGCATAGATGGGTCCATCGCTTTCTCTCAGGCGGGCTTTGTGATGAAA 8695
Qy 541 ACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTTGAAGAAAAAC 600
Db 8696 ACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTTGAAGAAAAAC 8755
Qy 601 AGAACACACGAGTTTTCAGTGGCTTCTTGAACCTCAATGATGTTTCCCATGAGGAGCCA 660
Db 8756 AGAACACACGAGTTTTCAGTGGCTTCTTGAACCTCAATGATGTTTCCCATGAGGAGCCA 8815
Qy 661 TTCAGCAAGTGGTTTGTGTCAGTAAATCTCAATATCCGTGGGACAGTACCCACCATTTTCGC 720
Db 8816 TTCAGCAAGTGGTTTGTGTCAGTAAATCTCAATATCCGTGGGACAGTACCCACCATTTTCGC 8875
Qy 721 TCCGCGACCGGCTACGTGTTTCTGCGGACGTCGCGAGTCAAGTGTACAATGTTCTTCCAAG 780
Db 8876 TCCGCGACCGGCTACGTGTTTCTGCGGACGTCGCGAGTCAAGTGTACAATGTTCTTCCAAG 8935
Qy 781 AGCGTCCCATCATTTAAATCTGAAGACGTGTTTGTGGGGCTCTGCTCGAAGGCTGAAC 840
Db 8936 AGCGTCCCATCATTTAAATCTGAAGACGTGTTTGTGGGGCTCTGCTCGAAGGCTGAAC 8995
Qy 841 ATCAGATTGGAGGAGCTCCACTCCAGCGGACGCTTTTTCAGGCGGGCTTACGCTTCTCC 900
Db 8996 ATCAGATTGGAGGAGCTCCACTCCAGCGGACGCTTTTTCAGGCGGGCTTACGCTTCTCC 9055
Qy 901 GTATGCTCTTTCAGGAGGATCGTGCCCTGCCACTTTCATCAAGGCTCGGAGCTCTCTTGAC 960
Db 9056 GTATGCTCTTTCAGGAGGATCGTGCCCTGCCACTTTCATCAAGGCTCGGAGCTCTCTTGAC 9115
Qy 961 TACTGGGAGGCTCTAGAGAAATCCCGGGGGAGAGTTGTGCGCCTGTCTGA 1011
Db 9116 TACTGGGAGGCTCTAGAGAAATCCCGGGGGAGAGTTGTGCGCCTGTCTGA 9166

RESULT 3
AF064860
LOCUS Homo sapiens chromosome 21 clone PAC 70124 map 21q22.3, complete
DEFINITION sequence.
ACCESSION AF064860
VERSION AF064860.2 GI:18958624
KEYWORDS HTG; HTGS DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 170121)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.-S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.-K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M.B., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Kawasaki,K., Asakawa,S.,
Shantani,A., Sasaki,T., Nagamine,K., Mitsuyma,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Riessmann,L., Dagand,E., Haaf,T., Wehrmeyer,S.,
Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H.,
Reinhardt,R. and Yaspo,M.Laure.
The DNA sequence of human chromosome 21
Nature 405 (6784), 311-319 (2000)
20289799
MEDLINE 10830953
REFERENCE 2 (bases 1 to 170121)
AUTHORS Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
```

```
TITLE Direct Submission
JOURNAL Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular
REFERENCE Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
AUTHORS 3 (bases 1 to 170121)
Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L., Rosenthal,A., Yaspo,M.-L. and
Rosenthal,A.
TITLE Direct Submission
JOURNAL Submitted (27-FEB-2002) Genome Analysis, Institute of Molecular
REFERENCE Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
AUTHORS 4 (bases 1 to 170121)
Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-2002) Genome Analysis, Institute of Molecular
COMMENT Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
FEATURES On Feb 27, 2002 this sequence version replaced gi:3171153.
source
1. 170121
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="PAC 70124"
ORIGIN
Query Match 99.5%; Score 1006.2; DB 9; Length 170121;
Best Local Similarity 99.7%; Pred. No. 2.1e-268;
Matches 1008; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 CCACCTCAGCCTCCTAGCATAAACTAGACACATCCTCATGCTTTTGGAGTCTAATCAATT 60
Db 89970 CCACCTCAGCCTCCTAGCATAAACTAGACACATCCTCATGCTTTTGGAGTCTAATCAATT 90029
Qy 61 GGATTTTGTCTTTCAGATGGCTTCCGGAAGATGAGATTGATGATATCTGCGCTCTG 120
Db 90030 GGATTTTGTCTTTCAGATGGCTTCCGGAAGATGAGATTGATGATATCTGCGCTCTG 90089
Qy 121 GTTCTGGGGGCTCTTGTGTTGTTATTTTAGCATGTACAGTCTAAATCTCTTCAAAGAACAG 180
Db 90090 GTTCTGGGGGCTCTTGTGTTGTTATTTTAGCATGTACAGTCTAAATCTCTTCAAAGAACAG 90149
Qy 181 TCCTTTGTTTACAAGAAAGACGGAACTTCCTTAAGCTCCAGATACAGACTCGAGGCGAG 240
Db 90150 TCCTTTGTTTACAAGAAAGACGGAACTTCCTTAAGCTCCAGATACAGACTCGAGGCGAG 90209
Qy 241 ACACCTCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
Db 90210 ACACCTCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 90269
Qy 301 GCCATCCGGCAGACGTGGGGGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
Db 90270 GCCATCCGGCAGACGTGGGGGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 90329
Qy 361 TTCCTCTCTGGGGGACACACAGCAGTGCAGCGGAAACAAAGAGGAGGAGGAGGAGGAGGAGGAG 420
Db 90330 TTCCTCTCTGGGGGACACACAGCAGTGCAGCGGAAACAAAGAGGAGGAGGAGGAGGAGGAGGAG 90389
Qy 421 CGACACGGGGACATATCCAGAGAGATTTCTAGACGCTTATTAATCAATCTGACCCCTGAAG 480
Db 90390 CGACACGGGGACATATCCAGAGAGATTTCTAGACGCTTATTAATCAATCTGACCCCTGAAG 90449
Qy 481 ACCATGATGGGCATAGATGGGTCCATCGCTTTTGTCTCTCAGGCGGGCTTTGTGATGAAA 540
Db 90450 ACCATGATGGGCATAGATGGGTCCATCGCTTTTGTCTCTCAGGCGGGCTTTGTGATGAAA 90509
Qy 541 ACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAATCTTCTTGAAGAAAAAC 600
Db 90510 ACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAATCTTCTTGAAGAAAAAC 90569
Qy 601 AGAACACACGAGTTTTCAGTGGCTTCTTGAACCTCAATGATGTTTCCCATGAGGAGCCA 660
```

Db 90570 AGAACCAACAGGTTTTTCTACTGGCTTCTTGAATCACTCAATGAGTTTCCCATCAGGCAGCCA 90629

Qy 661 TTCAGCAAGTGGTTTGTCTAGTAAATCTCAATATCGTGGGACAGTACCCACCACTTCTGC 720

Db 90630 TTCAGCAAGTGGTTTGTCTAGTAAATCTCAATATCGTGGGACAGTACCCACCACTTCTGC 90689

Qy 721 TCCGCGACCGGCTACGTTCTTTCTTGGCGACCTGGCGAGTCAGGTGTACAATGTCTCCAAG 780

Db 90690 TCCGCGACCGGCTACGTTCTTTCTTGGCGACCTGGCGAGTCAGGTGTACAATGTCTCCAAG 90749

Qy 781 AGCGTCCCATACATTAATACTGGAAGACGTGTTGTGGGGCTCTCGCTCGAAGGCTGAAC 840

Db 90750 AGCGTCCCATACATTAATACTGGAAGACGTGTTGTGGGGCTCTCGCTCGAAGGCTGAAC 90809

Qy 841 ATCAGATTGGAGGAGCTCCATCCCGACCGACCTTTTTCAGGGGGCTTACGTTTCTCC 900

Db 90810 ATCAGATTGGAGGAGCTCCATCCCGACCGACCTTTTTCAGGGGGCTTACGTTTCTCC 90869

Qy 901 GTATGCGCTCTTCAGGAGGATCGTGGCCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGAC 960

Db 90870 GTATGCGCTCTTCAGGAGGATCGTGGCCCTGCCACTTCATCAAGCCTCGGACTCTTGGAC 90929

Qy 961 TACTGGCAGGCTCTAGAGAATTCGCGGGGGAAGATTGTCCGCTGTCTGA 1011

Db 90930 TACTGGCAGGCTCTAGAGAATTCGCGGGGGAAGATTGTCCGCTGTCTGA 90980

RESULT 4

HS21C080 340000 bp DNA linear PRI 24-MAY-2000

LOCUS Homo sapiens chromosome 21 segment HS21C080.

DEFINITION AL163280 AP001735 BA000005

ACCESSION AL163280.2 GI:7717369

VERSION

KEYWORDS

SOURCE

ORGANISM Homo sapiens (human)

Homosapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 340000)

Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A., Zimmermann,W., Rosenthal,A., Kudon,J., Shibuya,K., Kawasaki,K., Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyma,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstieck,G., Hornischer,K., Brandt,P., Scharfe,H., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Rieselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Yaspo,M.L.

Direct Submission

Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing Consortium: \* RIKEN Genomic Sciences Center, Human Genome Research Group \* Institute of Molecular Biotechnology, Genome Analysis \* Keio University School of Medicine, Dept. of Molecular Biology \* GBF, Dept. of Genome Analysis \* Max-Planck Institute for Molecular Genetics (addresses see below)

The Chromosome 21 Mapping and Sequencing Consortium consists of

\* RIKEN Genomic Sciences Center, Human Genome Research Group, \* Sagami-hara 228-8555, Japan,

\* e-mail: sakaki@gs.c.riken.go.jp

\* URL: http://hgp.gsc.riken.go.jp/

and

\* Institute of Molecular Biotechnology, Genome Analysis, \* Beutenbergstrasse 11, D-07745 Jena, Germany,

\* e-mail: gscj-submit@genome.imb-jena.de

\* URL: http://genome.imb-jena.de/

and

\* Keio University School of Medicine, Dept. of Molecular Biology, \* Tokyo 160-8582, Japan,

\* e-mail: shimizu@dm-med.keio.ac.jp

\* URL: http://adenine.dmb.med.keio.ac.jp/

and

\* GBF, Dept. of Genome Analysis,

\* Mascheroder Weg 1, D-38124 Braunschweig, Germany, \* e-mail: info.genome@gbf.de

\* URL: http://genome.gbf.de/

and

\* Max-Planck Institute for Molecular Genetics, \* Ihnestrass 73, D-14195 Berlin, Germany,

\* e-mail: info-chr21@molgen.mpg.de

\* URL: http://chr21.rz-berlin.mpg.de/.

Location/Qualifiers

1. 340000

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="21"

/map="21q22.3"

<1..125946

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="21"

/map="21q22.3"

/clone="P160L9, 5' partial"

/clone\_lib="RPC11.3-5 PAC library"

/note="Accession No. AF121897"

73410..243533

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="21"

/map="21q22.3"

/clone="P70124"

/clone\_lib="RPC11.3-5 PAC library"

/note="Accession No. AF064860"

224137..>340000

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="21"

/map="21q22.3"

/clone="P206A10, 3' partial"

/clone\_lib="RPC11.3-5 PAC library"

/note="Accession No. AF121782"

286628..>340000

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="21"

/map="21q22.3"

/clone="BAC-291B3, 3' partial"

/clone\_lib="BAC library"

/note="Accession No. AF064857"

1088..1566

/note="LIM4"

/rpt\_family="LINE/L1"

/rpt\_type=DISPERSED

2398..2495

/notes="(TCTA)n"

/rpt\_family="Simple\_repeat"

/rpt\_type=TANDEM

complement(2496..2878)

/notes="THE1C"

/rpt\_family="LTR/MaLR"

/rpt\_type=DISPERSED

2879..2931

/notes="(TCTA)n"

/rpt\_family="Simple\_repeat"

/rpt\_type=TANDEM

2932..3004

/notes="(TCCA)n"

/rpt\_family="Simple\_repeat"

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat_region	/rpt_type=TANDEM 3005..3074 /note="(TCCA)n" /rpt_family="Simple_repeat" /rpt_type=TANDEM 3473..19157 /gene="SH3BGR" Join(3473..3565,12649..12678,15331..15460,18739..19157) /gene="SH3BGR" Join(3473..3565,12649..12678,15331..15426) /gene="SH3BGR" /note="Accession No. X93498" /codon_start=1 /product="21-Glutamic Acid Rich protein 21-GARP" /protein_id="CAB90445.1" /db_xref="GI:7717370" /db_xref="GOA:P55822" /db_xref="UniProt/Swiss-Prot:P55822" /translation="GSEKAEKGGTEAQKGEVDGNLPRAQKNEBEGETATETEE IAMEGAEEAEEREBETAEGERPGEDEDS" 3473..3565 /gene="SH3BGR" /number=4	
repeat_region	5542..5829 /note="AluSx" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(6887..7067) /note="L1MD3" /rpt_family="LINE/L1" /rpt_type=DISPERSED complement(8482..8762) /note="AluJo" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(8865..8964) /note="L2" /rpt_family="LINE/L2" /rpt_type=DISPERSED	
repeat_region	9131..9424 /note="AluSq" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(9723..9882) /note="L2" /rpt_family="LINE/L2" /rpt_type=DISPERSED complement(9989..10077) /note="L2" /rpt_family="LINE/L2" /rpt_type=DISPERSED complement(10078..10429) /note="THE1B" /rpt_family="LTR/MaLR" /rpt_type=DISPERSED complement(10430..10651) /note="L2" /rpt_family="LINE/L2" /rpt_type=DISPERSED 10980..11000 /note="(TTTA)n" /rpt_family="Simple_repeat" /rpt_type=TANDEM complement(11002..11282) /note="AluSp" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(11315..11556) /note="AluJo" /rpt_family="SINE/Alu" /rpt_type=DISPERSED 11702..11753 /note="MIR"	
gene	/rpt_type=TANDEM 3005..3074 /note="(TCCA)n" /rpt_family="Simple_repeat" /rpt_type=TANDEM 3473..19157 /gene="SH3BGR"	/rpt_family="SINE/MIR" /rpt_type=DISPERSED 12649..12678 /gene="SH3BGR" /number=5 complement(13115..13425) /note="AluSx" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(13598..13892)
repeat_region	10980..11000 /note="(TTTA)n" /rpt_family="Simple_repeat" /rpt_type=TANDEM complement(11002..11282) /note="AluSp" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(11315..11556) /note="AluJo" /rpt_family="SINE/Alu" /rpt_type=DISPERSED 11702..11753 /note="MIR"	Query Match 99.5%; Score 1006.2; DB 9; Length 340000; Best Local Similarity 99.7%; Pred. No. 2.2e-268; Matches 1008; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  QY 1 CCACCTCAGCCTCGTAGCATAAACTAGACACATCCTCATCTCTTTGAGGTCCTAATCATTT 60 DB 163378 CCACCTCAGCCTCGTAGCATAAACTAGACACATCCTCATCTCTTTGAGGTCCTAATCATTT 163437 QY 61 GGATTTTGTTCCTTTTCAGATGGCTTTCCCGAAGATGAGATTGATGTATATCTGCCCTTCG 120 DB 163438 GGATTTTGTTCCTTTTCAGATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCCTTCG 163497 QY 121 GTTCTGGGGGCTCTTTGTTGTATTTTAGCATGTACAGTCTAAATCTCTTTCAAAGAACAG 180 DB 163498 GTTCTGGGGGCTCTTTGTTGTATTTTAGCATGTACAGTCTAAATCTCTTTCAAAGAACAG 163557 QY 181 TCCTTTGTTTACAAGAAAGACGGCACTTCCTTAAGCTCCAGATACAGACTGCAGGGCAG 240 DB 163558 TCCTTTGTTTACAAGAAAGACGGCACTTCCTTAAGCTCCAGATACAGACTGCAGGGCAG 163617 QY 241 ACACCTCCCTTCCTCGCTCTGCTGACCTCATCCCAAAACAGTTGGCTGAGCGCATG 300 DB 163618 ACACCTCCCTTCCTCGCTCTGCTGACCTCATCCCAAAACAGTTGGCTGAGCGCATG 163677 QY 301 GCCATCCGGCAGACGTGGGGGAAAAGAGAGGACGGTGAAGGGAAGACAGCTGAAGACATTC 360 DB 163678 GCCATCCGGCAGACGTGGGGGAAAAGAGAGGATGGTGAAGGGAAGACAGCTGAAGACATTC 163737 QY 361 TTCTCTCTGGGGACCCAGCAGTGCAGCGGAAACAAAGAGGTGGACAGGAGGCCAG 420 DB 163738 TTCTCTCTGGGGACCCAGCAGTGCAGCGGAAACAAAGAGGTGGACAGGAGGCCAG 163797 QY 421 CGACACGGGGACATTTATCCAGAAGGATTTCTAGACGCTATTACAACTCTGACCCCTGAAG 480 DB 163798 CGACACGGGGACATTTATCCAGAAGGATTTCTAGACGCTATTACAACTCTGACCCCTGAAG 163857 QY 481 ACCATGATGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGGCGCGTTTGTGATGAAA 540 DB 163858 ACCATGATGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGGCGCGTTTGTGATGAAA 163917 QY 541 ACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACCTGCTCTCAAGAAAC 600 DB 163918 ACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACCTGCTCTCAAGAAAC 163977 QY 601 AGAACAAACAGGTTTTTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGCGACCA 660 DB 163978 AGAACAAACAGGTTTTTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGCGACCA 164037 QY 661 TTCAGCAAGTGGTTTGTTCAGTAAATCTGAAATATCCGTGGGACAGGTACCCACCATTTCTGC 720 DB 164038 TTCAGCAAGTGGTTTGTTCAGTAAATCTGAAATATCCGTGGGACAGGTACCCACCATTTCTGC 164097 QY 721 TCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAAGTGTACAATGTCTCCAAAG 780 DB 164098 TCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAAGTGTACAATGTCTCCAAAG 164157 QY 781 AGCGTCCCATACATTAATACTGGAAGAGCGTGTTCGTGGGGCTCTGCTCGAAGGCTGAAC 840 DB 164158 AGCGTCCCATACATTAATACTGGAAGAGCGTGTTCGTGGGGCTCTGCTCGAAGGCTGAAC 164217 QY 841 ATCAGATTGGAGGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGGCTTACGCTTCTTCC 900 DB 164217 ATCAGATTGGAGGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGGCTTACGCTTCTTCC 164277



```
QY 901 GTATCCCTCTTCAGAGGATCGTGCCTGCCCTCCACTTTCATCAAGCCTCGGACTCTCTTGGAC 960
Db 87493 GTATCCCGCTTCAGGAGGATCGTGCCTGCCCTCCACTTTCATCAAGCCTCGGACTCTCTTGGAC 87552

QY 961 TACTGGCAGGCTCTAGAGAAATCCCGGGGGGAGAGATTGTCGCCCTGTCTGA 1011
Db 87553 TACTGGCAGGCTCTAGAGAAATCCCGGGGGGAGAGATTGTCGCCCTGTCTGA 87603

RESULT 6
AB041416 1576 bp DNA linear PRI 17-OCT-2000
LOCUS Homo sapiens betal,3-Galt 5 gene for UDP-Gal:GlcNAc
DEFINITION betal,3-galactosyltransferase 5, partial cds.
ACCESSION AB041416
VERSION AB041416.1 GI:7593026
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1576)
AUTHORS Liu, Y. and Saitou, N.
TITLE Silver Project
JOURNAL Published Only in DataBase (2000)
REFERENCE 2 (bases 1 to 1576)
AUTHORS Liu, Y. and Saitou, N.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2000) Naruya Saitou, National Institute of
Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima,
Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp,
URL: http://saver.lab.nig.ac.jp/~silver/, Tel: 81-559-81-6790,
Fax: 81-559-81-6789)
FEATURES
Location/Qualifiers
1..1576
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="#056"
/db_xref="taxon:9606"
/note="human sequence used for primer design based on Acc#
AB020337"
<1..27
/number=3
exon 28..674
/number=3
intron 675..1576
/number=3
gene 675..1576
/gene="betal,3-Galt 5"
CDS 675..1576
/gene="betal,3-Galt 5"
/codon_start=1
/product="UDP-Gal:GlcNAc betal,3-galactosyltransferase 5"
/protein_id="BAA94501.1"
/db_xref="GI:7593027"
/translation="MAPEPKMLMYICLLVGLALCLYFSMYSINLPFKESFYVKDGNF
LKLPDTCRQTPPELILVTSISHQLAERMAIOTWGERIVKQKQLKTPFLGLGTS
AAETKEVDQSORHGDIIKDFLDVYVNLTKTWGLIEWVHRFCPOAFAVMTKTDPMF
INVDYDELLKXNRTPFPFGLKNEFPPIQPPKFWFSKSYPRDVPFPCSGTG
YVFGDVASQVYNKSVKPYIYKLDSDVFGVLCLEIRLBEHLHSQPTFFPGGLRFSVC
LFRIRIVAFHKPTLLDYWQALE"
675..1576
/gene="betal,3-Galt 5"
/number=4
exon

ORIGIN
Query Match 96.8%; Score 978.4; DB 9; Length 1576;
Best Local Similarity 99.9%; Pred. No. 9.4e-261;
Matches 979; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCACCTGAGCTCTAGCATAAACTAGACATCCTCATCTGCTTTGAGGCTCAATCATT 60
Db 597 CCACCTGAGCTCTAGCATAAACTAGACATCCTCATCTGCTTTGAGGCTCAATCATT 656
```

```
QY 61 GGATTTTGTCTCTTCCAGATGGCTTTCCGAGATGAGATTGATGATATATCTGCTTCTG 120
Db 657 GGATTTTGTCTCTTCCAGATGGCTTTCCGAGATGAGATTGATGATATATATCTGCTTCTG 716

QY 121 GTTCTGGGGGCTCTTTGTTGTTATTTTAGCATGTACAGTCTAAATCCCTTCAAAGACAG 180
Db 717 GTTCTGGGGGCTCTTTGTTGTTATTTTAGCATGTACAGTCTAAATCCCTTCAAAGACAG 776

QY 181 TCCTTTGTTTACAAGAAAGACGGGAACTTCCTTAAGCTCCAGATACAGACTGCGAGGCAG 240
Db 777 TCCTTTGTTTACAAGAAAGACGGGAACTTCCTTAAGCTCCAGATACAGACTGCGAGGCAG 836

QY 241 ACACCTCCCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
Db 837 ACACCTCCCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 896

QY 301 GCCATCCGGCAGAGCTGGGGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
Db 897 GCCATCCGGCAGAGCTGGGGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 956

QY 361 TTCTCTCTGGGGGACACCCAGCAGTGCAGCGGAAACAAAGAGAGGTGGACACGAGGAGCCAG 420
Db 957 TTCTCTCTGGGGGACACCCAGCAGTGCAGCGGAAACAAAGAGAGGTGGACACGAGGAGCCAG 1016

QY 421 CGACACGGGGACATATATCCAGAGAGATTTCTAGACGTCTATTACAATCTGACCTGAAG 480
Db 1017 CGACACGGGGACATATATCCAGAGAGATTTCTAGACGTCTATTACAATCTGACCTGAAG 1076

QY 481 ACCATGATGGGCATAGATGGTCCATCGCTTTTCTCTCAGGGCGGCTTGTGTATGAAA 540
Db 1077 ACCATGATGGGCATAGATGGTCCATCGCTTTTCTCTCAGGGCGGCTTGTGTATGAAA 1136

QY 541 ACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAAAAC 600
Db 1137 ACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAAAAC 1196

QY 601 AGAACAAACAGGTTTTTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGAGCCCA 660
Db 1197 AGAACAAACAGGTTTTTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGAGCCCA 1256

QY 661 TTCAGCAAGTGGTTTGTTCAGTAAATCTGAATATCGTGGGACAGGTACCCACCATCTTGC 720
Db 1257 TTCAGCAAGTGGTTTGTTCAGTAAATCTGAATATCGTGGGACAGGTACCCACCATCTTGC 1316

QY 721 TCCGGCACCGGCTACGTGTTTCTGGGACCTGGCGAGTCAAGTGTCAATGTTCTTCCCAAG 780
Db 1317 TCCGGCACCGGCTACGTGTTTCTGGGACCTGGCGAGTCAAGTGTCAATGTTCTTCCCAAG 1376

QY 781 AGCGTCCCATACATTAATCTGAAAGACGTGTTTGTGGGGCTCTGCTCGAAAGGCTGAAC 840
Db 1377 AGCGTCCCATACATTAATCTGAAAGACGTGTTTGTGGGGCTCTGCTCGAAAGGCTGAAC 1436

QY 841 ATCAGATTGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGGCTTAGCTTCTCC 900
Db 1437 ATCAGATTGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGGCTTAGCTTCTCC 1496

QY 901 GTATGCTCTTTCAGGAGATCGTGCCTGCGCACTTCATCAAGCCTCGGACTCTCTTGGAC 960
Db 1497 GTATGCTCTTTCAGGAGATCGTGCCTGCGCACTTCATCAAGCCTCGGACTCTCTTGGAC 1556

QY 961 TACTGGCAGGCTCTAGAGAA 980
Db 1557 TACTGGCAGGCTCTAGAGAA 1576

RESULT 7
AB041415 1570 bp DNA linear PRI 13-APR-2000
LOCUS Pan paniscus betal,3-Galt 5 gene for UDP-Gal:GlcNAc
DEFINITION betal,3-galactosyltransferase 5, partial cds.
ACCESSION AB041415
VERSION AB041415.1 GI:7593024
KEYWORDS
```





Qy	601	AGAACCAACAGGTTTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCA	660
Db	1197	AGAACCAACAGGTTTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCA	1256
Qy	661	TTCAGCAAGTGGTTTTCTCAGTAAATCTGAAATATCCGTGGGACAGGTACCCACCAATCTCGC	720
Db	1257	TTCAGCAAGTGGTTTTCTCAGTAAATCTGAAATATCCGTGGGACAGGTACCCACCAATCTCGC	1316
Qy	721	TCCGGCACCGGTACGTGTTTTCTGCGACGTCGGCGAGTCAGGTGTACAATGTCTCCAAG	780
Db	1317	TCCGGCACCGGTACGTGTTTTCTGCGACGTCGGCGAGTCAGGTGTACAATGTCTCCAAG	1376
Qy	781	AGCGTCCCATACATTTAACTGGAAGACGTGTTGTGGGGCTCTGCCTCGAAGGCTGGAAC	840
Db	1377	AGCGTCCCATACATTTAACTGGAAGACGTGTTGTGGGGCTCTGCCTCGAAGGCTGGAAC	1436
Qy	841	ATCAGATTGGAGGAGCTCCACTCCACGCGACCTTTTTTCCAGGGGCTTTACGCTTCTCC	900
Db	1437	ATCAGATTGGAGGAGCTCCACTCCAGCCGACCTTTTTTCCAGGGGCTTTACGCTTCTCC	1496
Qy	901	GTATGCCTCTTCAGGAGGATCGTGGCTGCGCACTTCATCAAGGCTCGGACTCTCTTGGAC	960
Db	1497	GTATGCCTCTTCAGGAGGATCGTGGCTGCGCACTTCATCAAGGCTCGGACTCTC-TGGAC	1555
Qy	961	TACTGGCAGG	970
Db	1556	TACTGGCAGG	1565

1	Qy	CGACCTCAGGCTCTCTAGCATAAACATAGACACATCCTCATGCTTTTGAGGTCTAAATCATTT	60
597	Db	CCACCTCAGGCTCTCTAGCATAAACATAGACACATCCTCATGCTTTTGAGGTCTAAATCATTT	656
61	Qy	GGATTTGTTCTCTTCAGATGGCTTTCCCGAAGATGAGATTGATGTATATCTGCCTTCGTG	120
657	Db	GGATTTGTTCTCTTCAGATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCTTCCTG	716
121	Qy	GTTCCTGGGGCTCTTTGTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAG	180
717	Db	GTTCCTGGGGCTCTTTGTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAG	776
181	Qy	TCCTTTGTTTACAAGAAAGAGGGAACTTCCTTAAGCTCCAGATACAGACTCGAGGCAG	240
777	Db	TCCTTTGTTTACAAGAAAGAGCGGAACTTCCTTAAGCTCCAGATACAGACTCGAGGCAG	836
241	Qy	ACACCTCCCTTCCTCGTCTGCTGGTGCACCTCATCCCAAAACAGTTGGCTGAGCGCATG	300
837	Db	ACACCTCCCTTCCTCGTCTGCTGGTGCACCTCATCCCAAAACAGTTGGCTGAGCGCATG	896
301	Qy	GCCATCCGGCAGACGTGGGGGAAAGAGAGGACCGGTGAAGGAAAGCAGCTGAAAGACATTC	360
897	Db	GCCATCCGGCAGACGTGGGGGAAAGAGAGGAGCGTGAAGGAAAGCAGCTGAAAGACATTC	956
361	Qy	TTCTCTCTGGGACCAACAGCAGTGCAGCGGAAACAAAGAGGTGGACCAAGAGAGGCAG	420
957	Db	TTCTCTCTGGGACCAACAGCAGTGCAGCGGAAACAAAGAGAGGTGGACCAAGAGAGGCAG	1016
421	Qy	CGACAGGGGACATTAATCCAGAGAGGATTTCCCTAGACGCTTATTAACAATCTGACCCCTGAAG	480
1017	Db	CGACAGGGGACATTAATCCAGAGAGGATTTCCCTAGACGCTTATTAACAATCTGACCCCTGAAG	1076
481	Qy	ACCATGATGGGCATAGAAATGGGTCATTCGCTTTTGTCTCCTCAGGCGGCGTTTGTGATGAAA	540
1077	Db	ACCATGATGGGCATAGAAATGGGTCATTCGCTTTTGTCTCCTCAGGCGGCGTTTGTGATGAAA	1136
541	Qy	ACAGACTCAGACATGTTTCATCAATGTTTGACTATCTGACTGAACTGCTTCGAGAGAAAAC	600
1137	Db	ACAGACTCAGACATGTTTCATCAATGTTTGACTATCTGACTGAACTGCTTCGAGAGAAAAC	1196

```

RESULT 9
AB041412
LOCUS
DEFINITION
AB041412
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
FEATURES
source
intron
gene
CDS

```

1360 bp DNA linear PRI 17-OCT-2000  
Gorilla gorilla betal,3-Galt 5 gene for UDP-Gal:GlcNAc  
betal,3-galactosyltransferase 5, partial cds.  
AB041412.1 GI:7593018  
Gorilla gorilla (gorilla)  
Gorilla gorilla  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.  
1 (bases 1 to 1360)  
Liu, Y. and Saitou, N.  
Silver Project  
Published Only in DataBase (2000)  
2 (bases 1 to 1360)  
Liu, Y. and Saitou, N.  
Direct Submission  
Submitted (11-APR-2000) Naruya Saitou, National Institute of  
Genetics, Laboratory of Evolutionary Genetics, 1111 Yata, Mishima,  
Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp,  
URL:http://saver.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790,  
Fax:81-559-81-6789)  
Location/Qualifiers  
1. 1360  
/organism="Gorilla gorilla"  
/mol\_type="genomic DNA"  
/isolate="#085"  
/db\_xref="taxon:9593"  
/notes="human sequence used for primer design based on Acc#  
AB020337"  
<1..466  
/number=3  
467..1360  
/gene="betal,3-Galt 5"  
467..>1360  
/gene="betal,3-Galt 5"  
/codon\_start=1  
/product="UDP-Gal:GlcNAc betal,3-galactosyltransferase 5"  
/protein\_id="BAA94497.1"  
/db\_xref="GI:7593019"  
/translaton="MAFPKRLMYICLLVLGALCLVFSMYSNLPFKEQSFVYKKDVF  
LKI.PDPCRTPELVLVITSSHKAERMAIROTQKERTVKGKOLKTFFLIGTTSS

exon  
 RAETKEVDQESRRGGDIIQKDFLDVYVNLTLKTMGIEWTHRRCPQAAAFVWKZDSDMF  
 INVDFITELLKKKNTTRFTFGFLKLNFEPPQFSSKWFVSKSGYPMDDRPVPPGCSGTG  
 YFSGDVAQVYNVSESVFYIILKEDVFGVLGRCLERLNRLBELHSQPTTFPPGGLRFSVC  
 RFRRIIVACHFIKPRTLDDYQA"  
 467. .>1360  
 /gene="beta1,3-GalT 5"  
 /number=4

Query Match	93.8%	Score 948.4	DB 9	Length 1360
Best Local Similarity	98.3%	Pred. No. 2.1e-252		
Matches 955	Conservative	3	Mismatches 14	Indels 0
QY	1	CCACCTCAGGCTCTCTAGCATAAACATAGACACATCCTCATGCTTTTCAGGTCTTAATCAT	60	
DB	389	CCACCTCAGGCTCTCTAGCATAAACATAGACACATCCTCATGCTTTTCAGGTCTTAATCAT	448	
QY	61	GGATTTTGTTCCTTTTCAGATGGCTTTCCCGAAGATGAGATTGATGTATATCTGCCTTCG	120	
DB	449	GGATTTTGTTCCTTTTCAGATGGCTTTCCCGAAGATGAGATTGATGTATATTCGCTTCG	508	
QY	121	GTTCGGGGCTCTTTGTTGTATTTAGCATGTAAGTCTTAAGCTCCAGATACAGACTCGAGCG	180	
DB	509	GTTCGGGGCTCTTTGTTGTATTTAGCATGTAAGTCTTAAGCTCCAGATACAGACTCGAGCG	568	
QY	181	TCCTTTGTTTACAGAAAGACGGGAACCTCCTTAAGCTCCAGATACAGACTCGAGCG	240	
DB	569	TCCTTTGTTTACAGAAAGACGGGAACCTCCTTAAGCTCCAGATACAGACTCGAGCG	628	
QY	241	ACACCTCCCTTCCTCGTCTCGTGGTGACCTCATCCCAAAACAGTTGGCTGAGCGCATG	300	
DB	629	ACACCTCCCTTCCTCGTCTCGTGGTGACCTCATCCCAAAACAGTTGGCTGAGCGCATG	688	
QY	301	GCATCCGGCAGACGTGGGGGAAAGAGAGACGCGTGAAGGAAAGCAGCTGAAGACATTC	360	
DB	689	GCCATCAGGCAGACGTGGGGGAAAGAGAGACGCGTGAAGGAAAGCAGCTGAAGACATTC	748	
QY	361	TTCCCTCTGGGGACCCACACGAGTGCAGCGGAACAAAGAGGTGGACAGGAGACCG	420	
DB	749	TTCCCTCTGGGGACCCACACGAGTGCAGCGGAACAAAGAGGTGGACAGGAGACCG	808	
QY	421	CGACACGGGACATTTATCCAGAGGATTTCTCTAGACGTCTATTACAATCTGACCCCTGA	480	
DB	809	CGACACGGGACATTTATCCAGAGGATTTCTCTAGAGTCTATTACAATCTGACCCCTGA	868	
QY	481	ACCATGATGGGCATAGAAATGGGTCATCGCTTTTGTCTCAGCGCGGTTGTGATGAAA	540	
DB	869	ACCATGATGGGCATAGAAATGGGTCATCGCTTTTGTCTCAGCGCGGTTGTGATGAAA	928	
QY	541	ACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAATGCTCTCGAAGAAAAC	600	
DB	929	ACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAATGCTCTCGAAGAAAAC	988	
QY	601	AGAACACACAGGTTTTTTCACCTGGCTCTTCAAACTCAATGAGTTTCCCATCAGCAGCCA	660	
DB	989	AGAACACACAGGTTTTTTCACCTGGCTCTTCAAACTCAATGAGTTTCCCATCAGCAGCCA	1048	
QY	661	TTCAGCAAGTGGTTTGTTCAGTAAATCTGAAATATCCGTGGGACAGGTACCCCATTCCTG	720	
DB	1049	TTCAGTAAAGTGGTTTGTTCAGTAAATCTGAAATATCCCATGGGACAGGTACCCCATTCCTG	1108	
QY	721	TCGGGCACCGGCTACGTTGTTTTCTGGGCGACGTGGCGAGTCAGGTGTACATGCTCCAAG	780	
DB	1109	TCGGGCACCGGCTACGTTGTTTTCTGGGCGACGTGGCGAGTCAGGTGTACATGCTCCAAG	1168	
QY	781	AGCGTCCCATACATTTAAACTGGGAAGACGTGTTTGTGGGGCTCTGCTCTCGAAGGCTCAAC	840	
DB	1169	AGCGTCCCATACATTTAAACTGGGAAGACGTGTTTGTGGGGCTCTGCTCTCGAAGGCTCAAC	1228	
QY	841	ATCAGATTGGAGAGCTCCACTCCACGCGACCTTTTTTTCAGGGGGCTTACGCTTCTCC	900	
DB	1229	ATCAGATTGGAGAGCTCCACTCCACGCGACCTTTTTTTCAGGGGGCTTACGCTTCTCC	1288	

Qy	901	GTATGCTCTTTCAGGAGGATCGTGGCTCGCACTTCATCAAGCTCGGACTCTCTTGGAC	960
Db	1289	GTATGCTCGCTTCAGGAGGATCGTGGCTCGCACTTCATCAAGCTCGGACTCTCTTGGAC	1348
Qy	961	TACTGGCAGGCT	972
Db	1349	TACTGGCAGGCT	1360
RESULT 10			
LOCUS	AB041414	1566 bp	DNA linear PRI 13-APR-2000
DEFINITION	Pan troglodytes betal.3-galt 5 gene for UDP-Gal:GlcNAc		
ACCESSION	AB041414	betal.3-galactosyltransferase 5, partial cds.	
VERSION	AB041414.1	GI:7593022	
KEYWORDS	Pan troglodytes (chimpanzee)		
SOURCE	Pan troglodytes		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.		
REFERENCE	1 (bases 1 to 1566)		
AUTHORS	Liu, Y. and Saitou, N.		
TITLE	Silver Project		
JOURNAL	Published Only in DataBase (2000)		
REFERENCE	2 (bases 1 to 1566)		
AUTHORS	Liu, Y. and Saitou, N.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-APR-2000) Naryu Saitou, National Institute of Genetics, Laboratory of Evolutionary Genetics; Iili Yata, Mishima, Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp, URL: http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790, Fax:81-559-81-6789)		
FEATURES	Location/Qualifiers		
source	1..1566		
	/organism="Pan troglodytes"		
	/mol_type="genomic DNA"		
	/isolate="chimp-202"		
	/db_xref="taxon:9598"		
	/note="human sequence used for primer design based on Acc# AB020337"		
exon	<1..28		
intron	29..675		
	/number=3		
gene	675..1566		
	/gene="betal.3-Galt 5"		
CDS	675..>1566		
	/gene="betal.3-Galt 5"		
	/codon_start=1		
	/product="UDP-Gal:GlcNAc betal,3-galactosyltransferase 5"		
	/protein_id="BA094499.1"		
	/db_xref="GI:7593023"		
	/translation="MAFFPKRMLMYCLLVGLALCVYFSMYSLNLFKEQSFVYKDGKFNFKLPTDCRQTPFPFVLLVTSRHRQLAERMAIQTWGKERTVKGQKLTFFLLGTSSD LAAETKEDVQSSQRGDIIOKDFLDVYVNLTKTMGIEWVHRCPOAAVFMKTDSDMF INVDYLTLLKKNRTPFTPTGLKNEFPFIRQPFKSWFVSKSEYKWDVYPPFCSGTG YFRRIVASVSYNNVSSPYIKEDVFVGLCLERLIRLELHSQTFPPFGGLRFSVCS YFRRIVACHFIKPRTLVDYMQ"		
exon	675..>1566		
	/gene="betal.3-Galt 5"		
	/number=4		
ORIGIN			
Query Match	93.5%	Score 945.6;	DB 9; Length 1566;
Best Local Similarity	98.4%	Pred. No. 1.3e-251;	
Matches 954; Conservative	1;	Mismatches 15;	Indels 0; Gaps 0;
Qy	1	CCACCTCAGCTCTCTAGCATAAACTAGACACATCCTCATGCTTTTCAGGCTCTAATCATT	60
Db	597	CCACCTCAGCTCTCTAGCATAAACTAGACACATCCTCATGCTTTTCAGGCTCTAATCATT	656
Oy	61	GGATTTTGTTCCTTTCAGTAGGGCTTTTCCCAGAAAGATGAGATTGATGTATATCTCGCTTCTG	120



```
Db 944 TCCACTCCAGCGACCTTTTTCAGGGGCTTACGCTTCTCGTAGCCTCTTCAGGA 1003
Qy 917 GGATCGTGGCGTGCACCTTCAAGCCCTCGGACTCTCTTGGACTACTTGGCAGGCTCTAG 976
Db 1004 GGATCGTGGCGTGCACCTTCAAGCCCTCGGACTCTCTTGGACTACTTGGCAGGCTCTAG 1063
Qy 977 AGAATTCCTCCGGGGGAAGATTGTCGCGCTGCTGA 1011
Db 1064 AGAATTCCTCCGGGGGAAGATTGTCGCGCTGCTGA 1098

RESULT 12
LOCUS HSA6078 933 bp DNA linear PRI 11-MAY-2000
DEFINITION Homo sapiens beta3gal-T6 gene.
ACCESSION AJ006078
VERSION beta-1.3-galactosyltransferase; beta3gal-T6 gene.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Amado,M., Carneiro,F. and Clausen,H.
TITLE Cloning and expression of two beta-1,3-galactosyltransferases:
beta3gal-T5 and beta3gal-T6
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 933)
AUTHORS Amado,M.
TITLE Direct Submission
JOURNAL Submitted (11-MAY-1998) Amado M., Department of Oral Diagnostics,
Royal dental School, Norre Alle 20, 2200 Copenhagen, DENMARK
FEATURES
source
1. .933
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
1. .933
/gene="beta3gal-T6"
1. .933
/gene="beta3gal-T6"
/codon_start=1
/product="beta-1,3-galactosyltransferase"
/protein_id="CAB91547.1"
/db_xref="GI:7799923"
/db_xref="GOA:Q9Y2C3"
/db_xref="UniProt/Swiss-Prot:Q9Y2C3"
/translation="MAPPKMLMVICLLVLGLCLXYSMNSLNPFPKQSFVYKXDNF
LKLPDTCROTPPLVLLVTSLSKQLAERMAIROTGWKERTVKGKQLTFELGTTSS
AAETKEVDQESQRHGDIIQKDFLDVYNNLTUKTMGIEWHRFCQAFAFMKTDSDMF
INDVLTLLKKRNTTFFGLKLNIEFFIRQPPKMFVSKSBYPMDRYPFPFCSGTG
YVFGSDVASQVYNSKVPYIKJEDVFVGLCLERLNRLIELSHQPTFFPFGGLRFSVC
LFRRIVACHFIKPRTLDDYMQALENSRGEDCPV"

Query Match 91.9%; Score 928.8; DB 9; Length 933;
Best Local Similarity 99.7%; Pred. No. 5.9e-24;
Matches 930; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 79 ATGGCTTTCCCGAAGATGAGATTGATGTATATCTGCCCTCTGTTCTGGGGCTCTTTGT 138
Db 1 ATGGCTTTCCCGAAGATGAGATTGATGTATATCTGCCCTCTGTTCTGGGGCTCTTTGT 60
Qy 139 TTGTATTTTACATGTACAGTCTAAATCCTTTCAAAGAACAGTCTTTGTTTACAAGAAA 198
Db 61 TTGTATTATAGCATGAACAGTCTAAATCCTTTCAAAGAACAGTCTTTGTTTACAAGAAA 120
Qy 199 GACGGGAATCTCTTAAGCTCCAGATACAGACTGCAGGAGACACCTCCCTTCTCGTC 258
Db 121 GACGGGAATCTCTTAAGCTCCAGATACAGACTGCAGGAGACACCTCCCTTCTCGTC 180
Qy 259 CTGCTGGTGACTCATCCCAACACAGTTGGTGTAGCGCATGGCCATCCGGCAGAGCTGG 318
```

```
Db 181 CTGCTGGTGACTCATCCCAAAACAGTTGGCTGAGGCGATGGCCATCCGGCAGAGCTGG 240
Qy 319 GGGAAAGAGAGGACGGTGAAGGGAAAGCAGCTGAAGACATTTCTTCTCTGGGACCAACC 378
Db 241 GGGAAAGAGAGGACGGTGAAGGGAAAGCAGCTGAAGACATTTCTTCTCTGGGACCAACC 300
Qy 379 AGCAGTGCAGCGGAAAACAAAGAGGTTGGACAGGAGCCAGCGACACGCGGACATTTATC 438
Db 301 AGCAGTGCAGCGGAAAACAAAGAGGTTGGACAGGAGCCAGCGACACATTTATC 360
Qy 439 CAGAAGGATTTCTTAGACGCTATTATACTGACCCCTGAAGACCATGATGGGCATAGAA 498
Db 361 CAGAAGGATTTCTTAGACGCTATTATACTGACCCCTGAAGACCATGATGGGCATAGAA 420
Qy 499 TGGGTCCATCGCTTTTGTCTCAGCGCGGTTTGTGATGAAAAACAGACTCAGACATGTTT 558
Db 421 TGGGTCCATCGCTTTTGTCTCAGCGCGGTTTGTGATGAAAAACAGACTCAGACATGTTT 480
Qy 559 ATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAACAGAAACAGAGTTTTC 618
Db 481 ATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAACAGAAACAGAGTTTTC 540
Qy 619 ACTGGCTTTTGAACCTCAATGAGTTTCCCCTCAGGCGCCATTCAGCAAGTGGTTTCTC 678
Db 541 ACTGGCTTTTGAACCTCAATGAGTTTCCCCTCAGGCGCCATTCAGCAAGTGGTTTCTC 600
Qy 679 AGTAAATCTGAATATCCGTGGGACAGGTACCCACATTTCTGCTCGGACCGGCTACGTG 738
Db 601 AGTAAATCTGAATATCCGTGGGACAGGTACCCACATTTCTGCTCGGACCGGCTACGTG 660
Qy 739 TTTTCTGGCGACGTGGCGAGTCTGACTGTAATCTCTCCAGAGCGTCCCATACATTTAAA 798
Db 661 TTTTCTGGCGACGTGGCGAGTCTGACTGTAATCTCTCCAGAGCGTCCCATACATTTAAA 720
Qy 799 CTGGAAGACGTGTTTGTGGGGCTCTGCTCTGAAAGGCTGAACATCAGATTGGAGGAGCTC 858
Db 721 CTGGAAGACGTGTTTGTGGGGCTCTGCTCTGAAAGGCTGAACATCAGATTGGAGGAGCTC 780
Qy 859 CACTCCAGCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCTCTTTCAGGAGG 918
Db 781 CACTCCAGCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCTCTTTCAGGAGG 840
Qy 919 ATCGTGGCTGCCACTTCATCAAGCCTCGGACTCTCTGGACTACTGCGAGGCTCTAGAG 978
Db 841 ATCGTGGCTGCCACTTCATCAAGCCTCGGACTCTCTGGACTACTGCGAGGCTCTAGAG 900
Qy 979 AATTCCCGGGGGGAAGATTGTCCGCTGTCTGA 1011
Db 901 AATTCCCGGGGGGAAGATTGTCCGCTGTCTGA 933

RESULT 13
LOCUS AB020337 2762 bp mRNA linear PRI 14-APR-2000
DEFINITION Homo sapiens mRNA for UDP-Gal:GlcNAc beta1,3-galactosyltransferase
5, complete cds.
ACCESSION AB020337
VERSION AB020337.1 GI:4835502
KEYWORDS UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Ieshiki,S., Togayachi,A., Kudo,T., Nishihara,S., Watanabe,M.,
Kubota,T., Kitajima,M., Shiraishi,N., Sasaki,K., Andoh,T. and
Narimatsu,H.
TITLE Cloning, expression, and characterization of a novel
UDP-galactose:beta-N-acetylglucosamine
beta1,3-galactosyltransferase (beta3gal-T5) responsible for
synthesis of type 1 chain in colorectal and pancreatic epithelia
and tumor cells derived therefrom
```

J. Biol. Chem. 274 (18), 12499-12507 (1999)

99230269

10212226

(bases 1 to 2762)

Ishiki, S., Togayachi, A. and Narimatsu, H.

Direct Submission

Submitted (20-NOV-1998) Hisashi Narimatsu, Soka University,

Institute of Life Science; 1-236, Tangi-cho, Hachioji, Tokyo

192-8577, Japan (E-mail: sieshiki@po.ijinet.or.jp,

Tel: 81-426-91-9466, Fax: 81-426-91-9315)

Location/Qualifiers

1. .2762

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/chromosome="21"

/map="21q22.3"

/cell\_lines="Colo 205"

/cell\_type="Adenocarcinoma"

1. .2762

/gene="betal,3-Galt 5"

1. .273

/gene="betal,3-Galt 5"

/number=1

274. .433

/gene="betal,3-Galt 5"

/number=3

434. .2762

/gene="betal,3-Galt 5"

/number=4

434. .1366

/gene="betal,3-Galt 5"

/codon\_start=1

/product="UDP-Gal:GlcNAc betal,3-galactosyltransferase 5"

/protein\_id="BAA77664.1"

/db\_xref="GI:4835503"

/translation="MAFPKRLMYICLLVLCALCLFMSYSLNPFKQSFVYKDGNF

LKLPDQCRQPPFLVLLVTSVSHKLAERMAIROTKERNVKGQKLTFFLLGTTSS

AAETKEVDQSGHDIQKDFLDVYVNLTKTWGIEWVHRFCQAAVFNKTDSDMF

INVLYTELLAKNRTFRFFGFLKNEFPPIROPSPKWFVSKSVPPDRYPFPCSGTG

YVFGDVASQVYNNYSKVPYIKLEDFVGLCLRLNRLLESLHSQTPFPFGLRPSVC

LFRRIVACHFKIPRLLDYLQALENSRGEDCPFV"

polyA\_site

/gene="betal,3-Galt 5"

/note="45 a nucleotides"

ORIGIN

Query Match 91.8%; Score 928.6; DB 9; Length 2762;

Best Local Similarity 99.6%; Pred. No. 7e-247;

Matches 931; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 77 AGATGGCTTCCGAGATGAGATTGATATCTGCTTCTGCTGGGGCTCTTT 136

DB 432 AAATGGCTTCCGAGATGAGATTGATATTTGCTTCTGCTGGGGCTCTTT 491

QY 137 GTTTGATTTTAGCATGTACAGTCTAAATCTTTCAAAGACAGTCTTTGTACAGA 196

DB 492 GTTTGATTTTAGCATGTACAGTCTAAATCTTTCAAAGACAGTCTTTGTACAGA 551

QY 197 AAGACGGGAACCTCTTAAGTCTCCAGATACAGACTGCAGGCAGACACTCCCTTCCTCG 256

DB 552 AAGACGGGAACCTCTTAAGTCTCCAGATACAGACTGCAGGCAGACACTCCCTTCCTCG 611

QY 257 TCCTGCTGCTGACCTCATCCACAAACAGTTGGCTGAGCGGATCGGCATCGGCAGAGT 316

DB 612 TCCTGCTGCTGACCTCATCCACAAACAGTTGGCTGAGCGGATCGGCATCGGCAGAGT 671

QY 317 GGGGAAAGAGAGAGCGGTGAAGGAAAGAGAGTCTTCTCTCTGGGGACCA 376

DB 672 GGGGAAAGAGAGAGTGTGAAGGAAAGAGAGTCTTCTCTCTGGGGACCA 731

QY 377 CCACAGTGCAGCGGAAACAAAGAGGTGGACCGAGAGCCAGCGACACGGGACATTA 436

Db 732 CCACAGTGCAGCGGAAACGAAAGAGGTGGACAGGAGAGCCAGGACACCGGGACATTA 791

QY 437 TCCAGAAAGGATTTCTAGACGCTCTATTACAACTTCGACCCCTGAAGACCATGATGGCATAG 496

Db 792 TCCAGAAAGGATTTCTAGACGCTCTATTACAACTTCGACCCCTGAAGACCATGATGGCATAG 851

QY 497 AATGGGTCCATCGCTTTTGTCTCCTCAGCGGGGCTTTGTGATGAAACACAGCTCAGACATGT 556

Db 852 AATGGGTCCATCGCTTTTGTCTCCTCAGCGGGGCTTTGTGATGAAACACAGCTCAGACATGT 911

QY 557 TCATCAATGTTGACTATCTGACTGAATCTGCTTCTGAAAGAAAACAGACAAACACAGGTTTT 616

Db 912 TCATCAATGTTGACTATCTGACTGAATCTGCTTCTGAAAGAAAACAGACAAACACAGGTTTT 971

QY 617 TCATCGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGACGCCATTTCAAGCAAGTGTGTTG 676

Db 972 TCATCGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGACGCCATTTCAAGCAAGTGTGTTG 1031

QY 677 TCAGTAAATCTGAATATCCGFGGACAGGTACCCACCAATTCGTCTCCGGCACCGGCTACG 736

Db 1032 TCAGTAAATCTGAATATCCGFGGACAGGTACCCACCAATTCGTCTCCGGCACCGGCTACG 1091

QY 737 TGTCTTCTGGCGAGCTGGCGAGTCAGGTGTACATGTCTCCAAGAGCGTCCCATACATTA 796

Db 1092 TGTCTTCTGGCGAGCTGGCGAGTCAGGTGTACATGTCTCCAAGAGCGTCCCATACATTA 1151

QY 797 AACTGGAAGACGTGTTGTGGGGCTCTGCCCTCGAAAGGCTCAACATCAGATTGGAGGAGC 856

Db 1152 AACTGGAAGACGTGTTGTGGGGCTCTGCCCTCGAAAGGCTCAACATCAGATTGGAGGAGC 1211

QY 857 TCCACTCCAGCCGACCTTTTCCAGGGGCTTTACGCTTCTCCGTATGCTCTTTCAGGA 916

Db 1212 TCCACTCCAGCCGACCTTTTCCAGGGGCTTTACGCTTCTCCGTATGCTCTTTCAGGA 1271

QY 917 GGATCGTGGCTGCCACTTCATCAAGCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG 976

Db 1272 GGATCGTGGCTGCCACTTCATCAAGCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG 1331

QY 977 AGAATTCGGGGGGAAGATTGTCGGCTGCTCGA 1011

Db 1332 AGAATTCGGGGGGAAGATTGTCGGCTGCTCGA 1366

RESULT 14

E38419

LOCUS

E38419

DEFINITION

Novel polypeptide.

ACCESSION

E38419

VERSION

E38419.1

GI:18626993

KEYWORDS

JP 2000245464-A/1.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 2775)

Novel polypeptide

Novel polypeptide

Patent: JP 2000245464-A 1 12-SEP-2000;

OS

PN

PD

PF

PI

PR

PC

CC



Qy	317	GGGGAAAGAGAGGACGCTGAAGGAAAGACAGCTGAAGACATTTCTTCCTCTGGGGACCA	376
Db	831	GGGGAAAGAGAGGATGCTGAAGGAAAGACAGCTGAAGACATTTCTTCCTCTGGGGACCA	890
Qy	377	CCAGCAGTGCAGCGAAGAAACAAAGAGGTGGACCAAGGAGCAGCAGCGGGACATTTA	436
Db	891	CCAGCAGTGCAGCGAAGAAACAAAGAGGTGGACCAAGGAGCAGCAGCGGGACATTTA	950
Qy	437	TCCAGAGGATTTCTCTAGACGCTCTATTACAACTCTGACCCCTGAAGACCAATGATGGGCATAG	496
Db	951	TCCAGAGGATTTCTCTAGACGCTCTATTACAACTCTGACCCCTGAAGACCAATGATGGGCATAG	1010
Qy	497	AATGGTCCATCGCTTTTCTCTCAGCGGGCTTTGTGATGAAACAGACTCAGACATGT	556
Db	1011	AATGGTCCATCGCTTTTCTCTCAGCGGGCTTTGTGATGAAACAGACTCAGACATGT	1070
Qy	557	TCATCAATGTTGACTATCTGACTGAACCTCTCTGAAGAAACAGAACCAACCAAGGTTTT	616
Db	1071	TCATCAATGTTGACTATCTGACTGAACCTCTCTGAAGAAACAGAACCAACCAAGGTTTT	1130
Qy	617	TCATCGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG	676
Db	1131	TCATCGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG	1190
Qy	677	TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCACCGGCTACG	736
Db	1191	TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCACCGGCTACG	1250
Qy	737	TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACATGTCTCCAAGAGCGTCCCATACATTTA	796
Db	1251	TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACATGTCTCCAAGAGCGTCCCATACATTTA	1310
Qy	797	AACTGGAAGACGTGTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC	856
Db	1311	AACTGGAAGACGTGTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC	1370
Qy	857	TCCACTCCCGACCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCTCTTTCAGGA	916
Db	1371	TCCACTCCCGACCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCTCTTTCAGGA	1430
Qy	917	GGATCGTGGCTTGCCTTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG	976
Db	1431	GGATCGTGGCTTGCCTTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG	1490
Qy	977	AGAATTTCCCGGGGGAAGATTGTCCGCTGTCTGA	1011
Db	1491	AGAATTTCCCGGGGGAAGATTGTCCGCTGTCTGA	1525

Search completed: September 22, 2005, 14:50:00  
Job time : 4630.4 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2005, 12:09:12 ; Search time 4457.86 Seconds  
(without alignments)  
8632.613 Million cell updates/sec

Title: US-10-777-828-8  
Perfect score: 1011  
Sequence: 1 ccactcagcctctagcat.....aagattgcgcctgtctga 1011

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gsa1:\*  
9: gb\_gsa2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	931.4	92.1	933	9 AY419460	AY419460 Homo sapi
2	757.2	74.9	780	9 AY419461	AY419461 Pan trogl
3	525.6	52.0	5069	3 BC047347	BC047347 Mus muscu
4	517.6	51.2	927	9 AY419462	AY419462 Mus muscu
5	457.4	45.2	766	9 CB132434	CB132434 tigr-ges-
6	369.6	36.6	434	2 BF748181	BF748181 RC3-BN041
7	327.6	32.4	338	1 AJ003597	AJ003597 AJ003597
8	284.8	28.2	492	5 BP761359	BP761359 BP761359
9	260.6	25.8	447	6 CB546569	CB546569 AMGNNUC:N
10	234	23.1	400	8 AZ768940	AZ768940 IM0569K09
11	185.6	18.4	1101	9 CNS05BT1	ALJ30238 Tetraodon
12	184.2	18.2	629	7 CR564113	CR564113 CR564113
13	182	18.0	981	9 AY401366	AY401366 Mus muscu
14	173	17.1	493	1 AJ507051	AJ507051 AJ507051
15	170	16.8	2705	3 AK035215	AK035215 Mus muscu
16	164	16.2	636	4 BG965086	BG965086 602829157
17	163.4	16.2	981	9 AY401364	AY401364 Homo sapi
18	161.4	16.0	981	9 AY401365	AY401365 Pan trogl
19	156.6	15.5	626	5 BX302370	BX302370 BX302370
20	154.8	15.3	901	4 B3740020	B3740020 B3740020
21	144.8	14.3	1903	3 AK088407	AK088407 Mus muscu
22	144.2	14.3	632	9 CC554157	CC554157 CH240_460
23	143.2	14.2	996	9 AY402383	AY402383 Mus muscu
24	141	13.9	546	7 CR628501	CR628501 DKF2p459G

25	139.2	13.8	685	5 BUI33295	BUI33295 603122104
26	138.6	13.7	994	7 CF585163	CF585163 AGENCOURT
27	137	13.6	1514	3 AK003837	AK003837 Mus muscu
28	136	13.5	810	7 CF749272	CF749272 UI-M-HJO-
29	133.6	13.2	996	9 AY402381	AY402381 Homo sapi
30	132.4	13.1	682	6 BY704097	BY704097 BY704097
31	132.2	13.1	596	4 BUI301651	BUI301651 BUI301651
32	132	13.1	3564	3 BC028571	BC028571 Homo sapi
33	127	12.6	233	2 BB595953	BB595953 BB595953
34	123.6	12.2	811	7 CO396097	CO396097 AGENCOURT
35	123.4	12.2	444	4 BM403751	BM403751 zam6155_Z
36	123.2	12.2	638	4 BUI616136	BUI616136 BUI616136
37	122.6	12.1	879	5 BQ895238	BQ895238 AGENCOURT
38	122	12.1	842	5 BUI50312	BUI50312 603527845
39	121.2	12.0	792	7 CK030647	CK030647 AGENCOURT
40	121.2	12.0	1238	9 AY420414	AY420414 Homo sapi
41	120.6	11.9	652	5 BUI627236	BUI627236 UI-H-FGO-
42	119.4	11.8	697	6 CB233493	CB233493 AGENCOURT
43	118.6	11.7	1021	9 CNS044T4	AL274513 Tetraodon
44	118	11.7	1238	9 AY420416	AY420416 Mus muscu
45	118	11.7	2681	3 AK036141	AK036141 Mus muscu

ALIGNMENTS

RESULT 1  
AY419460 933 bp DNA linear GSS 17-DEC-2003  
LOCUS Homo sapiens B3GALT5 gene, VIRTUAL TRANSCRIPT, partial sequence,  
DEFINITION genomic survey sequence.  
ACCESSION AY419460  
VERSION AY419460.1 GI:39775417  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 933)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarimal,A.,  
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trices  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 933)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarimal,A.,  
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT These sequences were made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
Location/Qualifiers  
source 1..933  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
<1..>933  
/gene="B3GALT5"  
/locus\_tag="HCM6898"  
ORIGIN  
Query Match 92.1%; Score 931.4; DB 9; Length 933;  
Best Local Similarity 99.9%; Pred. No. 1.2e-265;  
Matches 932; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 79 ATGCGTTTCCGAGATGAGATTCATATATATCGCTTCGTGGGGCTTTGT 138  
|||||

```

Db      1  ATGGCTTTCCCGAAGATGAGATTGATGTATATTTGGCTTCTGCTTCTGGGGCTCTTTGT 60
Qy      139 TTGTATTTTACGCTACAGCTAAATCTTTCAAGAACAGTCTTTGTTTCAAGAAA 198
Db      61  TTGTATTTTACGCTACAGCTAAATCTTTCAAGAACAGTCTTTGTTTCAAGAAA 120
Qy      199 GACGGAACTCTCTTAAGCTCCAGATACAGACTGACGAGACAGTCTCTCTCTCGTC 258
Db      121 GACGGAACTCTCTTAAGCTCCAGATACAGACTGACGAGACAGTCTCTCTCTCGTC 180
Qy      259 CTGCTGTGACTCATATCCCAACAGTTGCTGAGCGCATGGCCATCCGGCAGACGTGG 318
Db      181 CTGCTGTGACTCATATCCCAACAGTTGCTGAGCGCATGGCCATCCGGCAGACGTGG 240
Qy      319 GGGAAAGAGAGAGCGGTCAAGGAAAGAGCTGAAGACATTTCTCTCTCTGGGACCAACC 378
Db      241 GGGAAAGAGAGAGCGGTCAAGGAAAGAGCTGAAGACATTTCTCTCTCTGGGACCAACC 300
Qy      379 AGCAGTCAGCGGAAACAAAGAGGTGGAACAGAGAGCGGACGACACGCGGACATATC 438
Db      301 AGCAGTCAGCGGAAACAAAGAGGTGGAACAGAGAGCGGACGACACGCGGACATATC 360
Qy      439 CAGAAGGATTTCTCTGAGACGTCTATTACAATCTGACCTGAAAGACCATGATGGCATAGAA 498
Db      361 CAGAAGGATTTCTCTGAGACGTCTATTACAATCTGACCTGAAAGACCATGATGGCATAGAA 420
Qy      499 TGGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAACAGACTCAGACATGTTTC 558
Db      421 TGGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAACAGACTCAGACATGTTTC 480
Qy      559 ATCAATGTTGACTATCTGACTGAACTGCTTCTGATGAAACAGAAACAGAACTGTTTGC 618
Db      481 ATCAATGTTGACTATCTGACTGAACTGCTTCTGATGAAACAGAAACAGAACTGTTTGC 540
Qy      619 ACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGACGCGCATTCAGCAAGTGGTTGTC 678
Db      541 ACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGACGCGCATTCAGCAAGTGGTTGTC 600
Qy      679 AGTAAATCTGAATATCGTGGGACAGGTACCCACCATTTCTCTCTGGGACCGGCTACGTG 738
Db      601 AGTAAATCTGAATATCGTGGGACAGGTACCCACCATTTCTCTCTGGGACCGGCTACGTG 660
Qy      739 TTTTCTGGGACGCTGGGAGTCAGGTGACAAATGTTCTCAAGAGCGTCCCATACATATAA 798
Db      661 TTTTCTGGGACGCTGGGAGTCAGGTGACAAATGTTCTCAAGAGCGTCCCATACATATAA 720
Qy      799 CTGGAAGACGTGTTTGTGGGCTCTGCTCGAAAGGCTGAACATCAGATTGGAGGAGCTC 858
Db      721 CTGGAAGACGTGTTTGTGGGCTCTGCTCGAAAGGCTGAACATCAGATTGGAGGAGCTC 780
Qy      859 CACTCCGAGCGACCTTTTTCAGGGGGCTTACGCTTCTCGGTATGCTCTTTCAGGAGG 918
Db      781 CACTCCGAGCGACCTTTTTCAGGGGGCTTACGCTTCTCGGTATGCTCTTTCAGGAGG 840
Qy      919 ATCTGGGCTGCCACTTCAATCAAGCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAG 978
Db      841 ATCTGGGCTGCCACTTCAATCAAGCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAG 900
Qy      979 AATTCCCGGGGGGAGATTGTCCGCTCTCTGA 1011
Db      901 AATTCCCGGGGGGAGATTGTCCGCTCTCTGA 933

```

```

RESULT 2
AY419461
LOCUS      780 bp      DNA      linear      GSS 17-DEC-2003
DEFINITION Pan troglodytes B3GALT5 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION  AY419461
VERSION     AY419461.1  GI:39775418
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM    Pan troglodytes

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1 (bases 1 to 780)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
14671302
2 (bases 1 to 780)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES             Location/Qualifiers
     source            1..780
                        /organism="Pan troglodytes"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9598"
     gene              <1..>780
                        /gene="B3GALT5"
                        /locus_tag="HCM6898"
     ORIGIN

Query Match       74.9%;   Score 757.2;   DB 9;   Length 780;
Best Local Similarity 98.1%;   Pred. No. 9.4e-214;
Matches 765;   Conservative 0;   Mismatches 15;   Indels 0;   Gaps 0;

Qy      79  ATGGCTTTCCCGAAGATGAGATTGATGTATATTTGGCTTCTGCTTCTGGGGCTCTTTGT 138
Db      1  ATGGCTTTCCCGAAGATGAGATTGATGTATGTTTGGCTTCTGGGGCTCTTTGT 60
Qy      139 TTGTATTTTACGCTACAGCTAAATCTTTCAAGAACAGTCTCTTTGTTTCAAGAAA 198
Db      61  GTGTATTTTACGCTACAGCTGAAATCTTTTCAAGAACAGTCTTTGTTTCAAGAAA 120
Qy      199 GACGGAACTCTCTTAAGCTCCAGATACAGACTGACGAGGACAGACCTCTCTCTCGTC 258
Db      121 GACGGAACTCTCTTAAGCTCCAGATACAGACTGACGAGGACAGACCTCTCTCTCGTC 180
Qy      259 CTGCTGTGACTCATATCCCAACAGTTGCTGAGCGCATGGCCATCCGGCAGACGTGG 318
Db      181 CTGCTGTGACTCATATCCCAACAGTTGCTGAGCGCATGGCCATCCGGCAGACGTGG 240
Qy      319 GGGAAAGAGAGAGCGGTGAAGGAAAGAGCTGAAGACATTTCTCTCTGGGACCAACC 378
Db      241 GGGAAAGAGAGAGCGGTGAAGGAAAGAGCTGAAGACATTTCTCTCTGGGACCAACC 300
Qy      379 AGCAGTCAGCGGAAACAAAGAGGTGGAACAGAGAGCGGACGACACGCGGACATATC 438
Db      301 AGCAGTCAGCGGAAACAAAGAGGTGGAACAGAGAGCGGACGACACGCGGACATATC 360
Qy      439 CAGAAGGATTTCTCTGAGACGTCTATTACAATCTGACCTGAAAGACCATGATGGCATAGAA 498
Db      361 CAGAAGGATTTCTCTGAGACGTCTATTACAATCTGACCTGAAAGACCATGATGGCATAGAA 420
Qy      499 TGGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAACAGACTCAGACATGTTTC 558
Db      421 TGGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAACAGACTCAGACATGTTTC 480
Qy      559 ATCAATGTTGACTATCTGACTGAACTGCTTCTGATGAAACAGAAACAGAACTGTTTTC 618
Db      481 ATCAATGTTGACTATCTGACTGAACTGCTTCTGATGAAACAGAAACAGAACTGTTTTC 540
Qy      619 ACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGACGCGCATTCAGCAAGTGGTTGTC 678

```

Db 541 ACTGCTCTTCTGAAGTCAATGAGTTTCCCATCAGCGAGCCGATTCAGTAAGTGGTTTCTC 600  
 QY 679 AGTAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCGGACACGGCTACGTG 738  
 Db 601 AGTAATCTGAATATCCGTGGGACAGGTACCCGATTTCTGCTCGGACACGGCTACGTG 660  
 QY 739 TTTTCTGGGAGCGTGGGAGTCAGGTGTACAAATGTCTCAAGAGCGTCCCATACATTAATA 798  
 Db 661 TTTTCCGGCGAGCGTGGGAGTCAGGTGTACAAATGTCTCGAGAGCGTCCCATACATTAATA 720  
 QY 799 CTGGAAGACGCTGTTGTGGGCTCTGCTCGAAGGCTGAAATCAGATTCAGAGGAGCTC 858  
 Db 721 CTGGAAGACGCTGTTGTGGGCTCTGCTCGAAGGCTGAAATCAGATTCAGAGGAGCTC 780

## RESULT 3

BC047347  
 LOCUS BC047347 5069 bp mRNA linear HTC 02-JAN-2004  
 DEFINITION Mus musculus UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase,  
 polypeptide 5, mRNA (cdna clone IMAGE:5365404).

ACCESSION BC047347  
 VERSION BC047347.1 GI:28611171  
 KEYWORDS HTC.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 5069)  
 AUTHORS Straubeberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toehiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
 Wokley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whitling, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalusz, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

TITLE Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 PUBMED 12477932

REFERENCE 2 (bases 1 to 5069)  
 AUTHORS Straubeberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-FEB-2003) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: The Cepko Laboratory  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegged, H.,  
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 53 Row: j Column: 9  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 2212432  
 This clone has the following problem: no 5' EST match.

## FEATURES

Location/Qualifiers  
 source  
 1..5069  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5365404"  
 /tissue\_type="Eye, retina, mouse strain C57BL/6"  
 /clone\_lib="NIH MGC 94"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"

## ORIGIN

Query Match 52.0%; Score 525.6; DB 3; Length 5069;  
 Best Local Similarity 73.3%; Pred. No. 2e-144;  
 Matches 704; Conservative 0; Mismatches 244; Indels 12; Gaps 2;  
 QY 55 ATCATTTGGATTTTCTTCCTTTTCAGATGGCTTTCCGGAAGATGAGATTGATGTATATCTGC 114  
 Db 568 AACGTGAGCTCTTACTCGGCAAGATGCTCACATGAAGACAAGCTAGTTTACGCCTCC 627  
 QY 115 CTCTCTGGTTCTGGGGGCTCTTTTGTGTATTTTATAGCATGTACAGTCTAAATCCTTTCAAA 174  
 Db 628 ATTCTCATGATGGCGCACTCTGCTGTGTACTTCAGCATGG-----ATCTTTTCAGA 678  
 QY 175 GAACAGTCTTTTGTGTACAGAAG--ACGGGAATCTTCTTAAGTCCCGAGATACAGAC 231  
 Db 679 GAATCCCGTTTGTGTATTTTAAAGAAAGTACGCGGAAGTTCTTCAGATTCCGGATATAGAC 738  
 QY 232 TGCAGGAGACACCTCCCTCTCTGCTGTGTGACTCATCCCAAAACAGTTTGGCT 291  
 Db 739 TGAAGCAGAAGCGCCCTTCTGCTGCTGTGTGAGCTCATCTCACAGCAGCTGGCC 798  
 QY 292 GAGCGCATGGCCATCCCGGACAGCTGGGGGAAAGAGAGGACGGTGAAGGGAAGACGAGCTG 351  
 Db 799 GCTCGCATGGCCATCCCGAAGACGCTGGGGTAGAGAGACATCTGTGCAGGGCCACAGGTG 858  
 QY 352 AAGACATTTCTTCTCGGGACACACAGCAGTGCAGCGGAAACAAAAGAGGTGAGACAG 411  
 Db 859 AGGACCTTCTTCTTCTGGGACCTCCGACAGCACCAGGAGATGGACGCCAACCCCTG 918  
 QY 412 GAGAGCCAGACACGGGACATTTATCCAGAGGATTTCTTAGAGCTTATTAACAATCTG 471  
 Db 919 GAGAGCGAGACAGCACCAGCATTTATCCAGAGGACTTCAAGGATGCTTACTTTCAACCTG 978  
 QY 472 ACCCTGAAGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCTCAGCGCGCGTTT 531  
 Db 979 ACCCTGAAGACCATGATGGGTATGGAATGGGTCTACCACTTTTGTCTCAGACAGCTTAC 1038  
 QY 532 GTGATGAAAACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACCTGTTCTG 591  
 Db 1039 GTGATGAAAACGGACTCTGACATGTTTGTGAATGTTGGCTATCTGACCGGAACCTGCTGCTA 1098  
 QY 592 AAGAAAACAGAACCAACCGGTTTTCCTGCTGGCTTTTGAACCTCAATGAGTTTCCCATC 651  
 Db 1099 AAGAAAACAAAACGACAGGTTCTTCACAGGCTACATAAAGCCGCCAGACTTTTCCCATC 1158  
 QY 652 AGGCGAGCATTTACGCAAGTGGTTTGTAGTAAATCTCGTGGGAGGAGGTACCCCA 711  
 Db 1159 CGGCAAGTTTCAACAGTGGTTTGTGAGTAAGTTTCGNAATATCTCTGGGACAGGTACCCA 1218  
 QY 712 CCATTCTGCTCCGGACCGGCTACGTTTCTTGGCGAGCTGGGAGTCAAGGTGTACAAT 771  
 Db 1219 CTTTCTGCTCCGCTACTGTTATGTTCTTTCCAGCGACGCTGGGATCCAAAGTATACAAC 1278  
 QY 772 GTCTCCAGAGCGTCCCATACATTAATACTGAAAGACGTTTGTGGGCTCTGCTCGAA 831  
 Db 1279 GTCTCAGAGAGCGTTCGCTTCATCAAGCTGAGATGTGTTTGTGGGCTCTGCTGGCC 1338

Qy	832	AGGCTGAACATCAGATTGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTAA	891
Db	1339	AAGCTTAAAGATCCGGCCGGAGGAGCTGCACACCAAAACAGACCTTCTTCCCTGGCGGTTTA	1398
Qy	892	CGTTCTCCCTATGCTCCTCTTCAGAGAGATCGTGGCCTGCCACTTCATCAAGCCTCGGACT	951
Db	1399	CGCTTTTCCGTGTGCGCGCTTTCAGAAATTTGTGCGATGCCATTTATTAAGAGCCCGAGGAC	1458
Qy	952	CTCTTGACTACTGGCAGGCTCTAGAGAAATTCGGGGGGAAGATTTCGCGCCTGTCTGA	1011
Db	1459	CTGCTCACTTACTGGCAAGCACTGGAGAACTCGAAGAACAGGACTGCTCTGTCTGCTGA	1518
RESULT 4			
AY419462			
LOCUS			
DEFINITION	AY419462	927 bp DNA linear GSS 17-DEC-2003	
ACCESSION	AY419462	Mus musculus B3GALT5 gene, VIRTUAL TRANSCRIPT, partial sequence,	
KEYWORDS	AY419462.1	genomic survey sequence.	
SOURCE	GSS.	GI:39775419	
ORGANISM	Mus musculus (house mouse)		
REFERENCE			
AUTHORS			
TITLE			
JOURNAL	Science 302 (5652), 1960-1963 (2003)		
PUBLISHED	14671302		
REFERENCE	2 (bases 1 to 927)		
AUTHORS	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.		
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios		
JOURNAL	Science 302 (5652), 1960-1963 (2003)		
PUBLISHED	14671302		
REFERENCE	2 (bases 1 to 927)		
AUTHORS	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA		
COMMENT	These sequences were made by sequencing genomic exons and ordering them based on alignment.		
FEATURES			
source	1..927		
	/organism="Mus musculus"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:10090"		
gene	<1..->927		
	/gene="B3GALT5"		
	/locus_tag="HCM6898"		
ORIGIN			
Query Match	51.2%	Score 517.6; DB 9; Length 927;	
Best Local Similarity	73.7%	Pred. No. 2.5e-142;	
Matches 690; Conservative	0; Mismatches 234; Indels 12; Gaps 2;		
Qy	79	ATGCTTTCCCGAAGATGAGATTGATATATCTGCTTCTGGTTCGCGGGCTCTTTGT	138
Db	1	ATGCTTCACATGAAGACAAGGCTAGTTTACGCTCCATTCTCATGATGGCGCACTGCG	60
Qy	139	TTGCTATTTTACGATGTACAGTCTTAATCCCTTTCAAAGAACAGTCTTGTGTACAGAAA	198
Db	61	TTGTACTTTCAGCATGG-----ATTCTTTCAGAGAACTCCCGTTGTTTAAAGAAA	111
Qy	199	G----ACGGGAACCTTCCTTAAGCTCCCAAGATACAGACTGCAGGCAGACACCTCCCTCCCTC	255
Db	112	AGTCACGGGAAGTTCCTTCAGATTCGGATATAGACTGCAGCAGAGACCGCTTCTCTG	171
Qy	256	GTCTGTCTGTGACCTCATCCCAACAAACAGTTGGCTGAGCGCATGCGCAATCCGGCAGACG	315
Db	172	GTGCTGTGCTGACGTCACTCAACAGACGTGGCCGCTCGCATGGCCATCCGCAAGACG	231

QY	316	TGGGGGAAAGAGAGGACGCGTGAAGGGGAAAGCAGCTGAAGACATTTCTTCTCTCTCTGGGGACC	375
Db	232	TGGGGTAGAGAGACATCTGTGCGAGGGCCAAACAGGTGAGGACCTTTCTTCTTCTGGGGACC	291
QY	376	ACCAGCAGTCCAGCGGAAACAAAGAGGTTGGACCCAGGAGGCCAGCGACACGGGACACATT	435
Db	292	TCCGACAGCACCGGAGGATGGAGCGCCACAACCCCTGGAGAGCGAGCAGCATCCGCGACATT	351
QY	436	ATCCAGAAGGATTTCTTAGACGCTTATTACAATCTGACCCCTGAAGAGCATTGATGGGCATA	495
Db	352	ATCCAGAAGGATTTCAAGATGCTCTTCAACTGACCCCTGAAGACCATGATGGGTATG	411
QY	496	GAATGGGTCCATCGCTTTTGTCTCAGCGGGCGTTTGTGATGAAAAACAGATCTAGACATG	555
Db	412	GAATGGGTCTACCACTTTTGTCTCAGACAGCTTACGTTGATGAAAAACGGAATCTGACATG	471
QY	556	TTTCATCAATCTGACATCTGACCTGACCTCTTCTGAAGAAAAACAGAACCAACGAGTTT	615
Db	472	TTTGTGAATGTTGGCTATCTGACGGAATCTGCTTAAAGAAAAACAAACGACCAAGGTTT	531
QY	616	TTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGCAGCCATTGACGAAGTGGTTT	675
Db	532	TTCAAGGCTTACATTAAGCCCGACGCTTCCCATCCGCGAGAAGTTCAACAAGTGGTTT	591
QY	676	GTCAAGTAATCTGAATATCCGTGGGACAGGTACCCACCATCTCTGTCGCGCACCGGCTAC	735
Db	592	GTGAGTAAGTTCAATATCCCTGGGACAGGTACCCACCTTTTGTCTGTGACTGTTTAT	651
QY	736	GTGTTTTCTGGCGACGCGGCGAGTCAGGTGTACAAATGTCTCAAGAGCGTCCCATACATT	795
Db	652	GTCTTTTCCAGCGACGTGGCGATCCAAGTATCAACGCTCTCAGAGAGCGTCCGTTCACT	711
QY	796	AAACTGGAACACGTGTTTGTGGGCTCTGCTCGAAAGGCTGAACATCAGATTGGAGGAG	855
Db	712	AAGCTGAGGATGTGTTTGTGGGCTCTGCTGTCGCAAGCTTAAAGATCCGCGCGAGGAG	771
QY	856	CTCCTCCAGCGGACGCTTTTTCAGGGGGCTTACGCTTCTCGTATGCTCTTTCAGG	915
Db	772	CTGCACACCAACAGACCTTTCTTCCCTGGCGGTTTACGCTTTTCCGTTGCGGCTTTCAG	831
QY	916	AGGATGTGCGCTGCCACTTTCATCAAGCCTCGGAGCTCTCTGGACTACTGGCAGGCTCTA	975
Db	832	AAATTTGGCATGCCATTTTATGAAGCCCGCAGGACCTGCTCACTTACTGCGACGACTG	891
QY	976	GAGAAATCCCGGGGGAGATTGTCGCCCTGTCTGA	1011
Db	892	GAGAACTCGAAAGAAAGAGGACTGCGCTGCTCTGA	927
RESULT 5			
CE132434/c			
LOCUS			
DEFINITION			
tigr-gss-dog-17000326239815 Dog Library Canis familiaris genomic,			
genomic survey sequence.			
CE132434			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Canis familiaris (dog)			
Canis familiaris			
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
1. (bases 1 to 766)			
Kirkness, E.F., Batfai, V., Halpern, A.L., Levy, S., Remington, K.,			
Rusch, D.B., Deichev, A.L., Pop, M., Wang, W., Fraser, C.M. and			
Venter, J.C.			

RESULT 5	CE132434/c	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	COMMENT
----------	------------	-------	------------	-----------	---------	----------	--------	----------	-----------	---------	-------	---------	---------	--------	---------

CE132434 766 bp DNA linear GSS 25-SEP-2003  
tigr-gss-dog-17000326239815 Dog Library Canis familiaris genomic,  
genomic survey sequence.  
CE132434  
CE132434.1 GI:35233692  
GSS.  
Canis familiaris (dog)  
Canis familiaris  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
1 (bases 1 to 766)  
Kirkness, E.F., Batina, V., Halpern, A.L., Levy, S., Remington, K.,  
Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and  
Venter, J.C.  
The dog genome: survey sequencing and comparative analysis  
Science 301 (5641), 1898-1903 (2003)  
22875432  
14512627  
Contact: Kirkness EF  
The Institute for Genomic Research  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,

Qy	636	CAATGAGTTTCCCATCAGCAGCCATTTCAGCAAGTGGTTTGTCTAGTAAATCTCAATATCC	695
Db	434	CAATGAGTTTCCCATCAGCAGCCATTTCAGCAAGTGGTTTGTCTAGTAAATCTCAATATCC	375
Qy	696	GTGGGACAGGTACCCACCACTCTGCTCGGGCACCGGTACGTGTTTTCTGGCGACGTGC	755
Db	374	GTGGGACAGGTACCCACCACTCTGCTCGGGCACCGGTACGTGTTTTCTGGCGACGTGC	315
Qy	756	GAGTCAGGTGTACAATGTCCTCAAGAGCGTCCCATACATTAACTGGAAGACGTGTTGT	815
Db	314	GAGTCAGGTGTACCAATGTCCTCAAGAGCGTCCCATACATTAACTGGAAGACGTGTTGT	255
Qy	816	GGGGCTCTGCCCTCGAAAGGCTGAAATCATAGATTGGAGGAGCTCCACTCCCGACCGACCTT	875
Db	254	GGGGCTCTGCCCTCGAAAGGCTGAAATCATAGATTGGAGGAGCTCCACTCCCGACCGACCTT	195

```

Qy 876 TTTTCCAGGGGCTTACGCTTCTCGTATGCTCTTTCAGAGGATCGTGGCTGCCACTT 935
Db 194 TTTTCCAGGGGCTTACGCTTCTCGTATGCTCTTTCAGAGGATCGTGGCTGCCACTT 135
Qy 936 CATCAAGCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAGAATCCCGGGGGGAAGA 995
Db 134 CATCAAGCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAGAATCCCGGGGGGAAGA 75
Qy 996 TTGTCCGCTGTCTGA 1011
Db 74 TTGTCCGCTGTCTGA 59

RESULT 7
AJ003597/c AJ003597 338 bp mRNA linear EST 04-DEC-1997
LOCUS AJ003597 Selected chromosome 21 cDNA library Homo sapiens cDNA
DEFINITION clone MPIp19-12J9, mRNA sequence.
ACCESSION AJ003597
VERSION AJ003597.1 GI:2578270
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 338)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 338)
Takeda, J., Ohara, O. and Seino, S.
Construction of a multi-functional cDNA library specific for mouse
pancreatic islets and its application to microarray
Unpublished (2004)
JOURNAL Contact: Susumu Seino
COMMENT Division of Cellular and Molecular Medicine
Kobe University Graduate School of Medicine
7-5-1 Kusunoki-cho, Chuo-ku, Kobe, Hyogo 650-0017, Japan
Tel: 81-78-382-5360
Fax: 81-78-382-5370
Email: seino@med.kobe-u.ac.jp.
FEATURES
Location/Qualifiers
1..338
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="21q"
/clone="MPIp19-12J9"
/clone_lib="Selected chromosome 21 cDNA library"
ORIGIN
Query Match 32.4%; Score 327.6; DB 1; Length 338;
Best Local Similarity 98.5%; Pred. No. 7e-86;
Matches 330; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 93 GATGAGTTGATGTATATCTGCTTCTGTTCTGGGGCTCTTTGTTGTATTTAGCAT 152
Db 335 GACGAGTTGATGTATATTTGTTCTGGTTCTGGGGCTCTTTGTTGTATTTAGCAT 276
Qy 153 GTACAGTCTAAATCTTTCAAAGAACAGTCTCTTTGTTTACAAGAAAGACGGAACTTCT 212
Db 275 GTACAGTCTAAATCTTTCAAAGAACAGTCTCTTTGTTTACAAGAAAGACGGAACTTCT 216
Qy 213 TAAGTCCAGATACAGACTGAGCAGACAGCTCCCTCTCTGCTCTGCTGTTGACCTC 272
Db 215 TAAGTCCAGATACAGACTGAGCAGACAGCTCCCTCTCTGCTCTGCTGTTGACCTC 156
Qy 273 ATCCCAAAACAGTTGGCTGAGCGATCGGCATCGGGCAGAGCTGGGGGAAAGAGAGGAC 332
Db 155 ATCCCAAAACAGTTGGCTGAGCGATCGGCATCGGGCAGAGCTGGGGGAAAGAGAGGAC 96
Qy 333 GGTGAAGGGAAGCAGCTGAAGACATTTCTTCTCTGGGGGACCCAGCAGCTGCAGCGGA 392
Db 95 GGTGAAGGGAAGCAGCTGAAGACATTTCTTCTCTGGGGGACCCAGCAGCTGCAGCGGA 36
Qy 393 AACAAAAGAGGTGGACAGGAGAGCCAGGACAGC 427
Db 35 AACAAAAGAGGTGGACAGGAGAGCCAGGACAGC 1

```

```

RESULT 8
BP761359 492 bp mRNA linear EST 10-JUL-2004
LOCUS BP761359 mouse (C57BL/6) pancreatic islet library with
DEFINITION recombination-based method Mus musculus cDNA clone mic03025 5',
mRNA sequence.
ACCESSION BP761359
VERSION BP761359.1 GI:50220057
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 492)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 492)
Nishimura, M., Yokoi, N., Miki, T., Horikawa, Y., Yoshioka, H.,
Takeda, J., Ohara, O. and Seino, S.
Construction of a multi-functional cDNA library specific for mouse
pancreatic islets and its application to microarray
Unpublished (2004)
JOURNAL Contact: Susumu Seino
COMMENT Division of Cellular and Molecular Medicine
Kobe University Graduate School of Medicine
7-5-1 Kusunoki-cho, Chuo-ku, Kobe, Hyogo 650-0017, Japan
Tel: 81-78-382-5360
Fax: 81-78-382-5370
Email: seino@med.kobe-u.ac.jp.
FEATURES
Location/Qualifiers
1..492
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="mic03025"
/sex="male"
/tissue type="pancreatic islet"
/dev stage="adult"
/clone_lib="mouse (C57BL/6) pancreatic islet library with
recombination-based method"
ORIGIN
Query Match 28.2%; Score 284.8; DB 5; Length 492;
Best Local Similarity 76.5%; Pred. No. 4.6e-73;
Matches 349; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
Qy 201 CGGAACCTTCCTTAAGTCCAGATACAGACTGAGCAGGACACCTCCCTTCCTGCTCT 260
Db 24 CGGAAGTTCCTTCAGATTCGGATATAGACTGCAAGCAGAAAGCCGCTTTCTCTGGTGT 83
Qy 261 GCTGCTGACCTCATCCCAACAAACAGTTGGCTGAGCGCATGGCCATCGGGCAGAGCTGGG 320
Db 84 GCTGCTGACCTCATCTCAAGCAGCTGGCGCTCGCATGGCCATCGCAAGAGCTGGGG 143
Qy 321 GAAAGAGAGGACGGTGAAGGAAAGCAGCTGAAGACATTTCTCTCTGGGGACCCAG 380
Db 144 TAGAGAGACATCTCTGAGGGCCAAACAGTGAAGTCTCTCTCTCTCTCTCTCTCTCGA 203
Qy 381 CAGTGCAGCGAAACAAAGAGGTGGACAGGAGCCAGGACGAGCAGCCGGGACATATATCA 440
Db 204 CAGCAGCAGGAGATGGACGCGCAACCTCTGAGAGGAGCAGCAGCCGACATATATCA 263
Qy 441 GAAGGATTTCTAGACGTCTATTACATCTGACCTGAAGACCATGATGGGCATAGAAATG 500
Db 264 GAAGGACTTCAGAGTGCCTACTTCACTGACCTGAAACCATGATGGGTATGGAAATG 323
Qy 501 GGTCCATCGCTTTTGTCTCAGCGGGGCTTTGTGATGAAAAACAGACTCAGACATGTTTCA 560
Db 324 GGTCTACCATTTTGTCTCAGACAGCTTACGTGATGAAAAACGAGACTCTGACATGTTTGT 383
Qy 561 CAATGTTGATCTGACTGACTGCTTCTGAAAGAAAAACAGAACCAACAGGTTTTTTCAC 620
Db 384 GAATGTTGGCTTTCTGACCGAACTGCTGTTAAGAAAAACCAACGACAGGTTCTTTCAC 443
Qy 621 TGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCA 656

```

```
|||||
444 AGGCTACATTAAAGCCCAAGCACTTTTCCATTCGGCA 479

RESULT 9
CB546569
LOCUS CB546569 447 bp mRNA linear EST 01-APR-2003
DEFINITION AMGNNUC:NRHYS-00333-H3-A W Rat hypothalamus (10471) Rattus
norvegicus cDNA clone nrhys-00333-h3 5', mRNA sequence.
ACCESSION CB546569
VERSION CB546569
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 447)
AUTHORS Amgen EST Program.
JOURNAL Amgen Rat EST Program
COMMENT Unpublished (2003)
Contact: Dan Fitzpatrick
Amgen, Inc
One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00333 row: h column: 3.
Location/Qualifiers
1. 447
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="nrhys-00333-h3"
/clone_lib="W Rat hypothalamus (10471)"
/note="Vector: pSPOR1; Site 1: SalI; Site 2: NotI; W Rat
hypothalamus adult female Wistar rat avg. Insert size 2.3
Kb fraction 6 and 7"

FEATURES
source
1. 447
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="nrhys-00333-h3"
/clone_lib="W Rat hypothalamus (10471)"
/note="Vector: pSPOR1; Site 1: SalI; Site 2: NotI; W Rat
hypothalamus adult female Wistar rat avg. Insert size 2.3
Kb fraction 6 and 7"

ORIGIN
Query Match 25.8%; Score 260.6; DB 6; Length 447;
Best Local Similarity 76.6%; Pred. No. 7,3e-66;
Matches 333; Conservative 0; Mismatches 99; Indels 3; Gaps 1;

QY 156 CAGTCTAAATCCTTTCAAGAACAGTCTCTTTGTTTACAGAAA---GACGGGAAGTTCCT 212
DB 13 CAGCATGAGTCTTTTCAAGAACTCCCTTTGTTCTTCAAGAAAGTCAAGGAAGTTCCT 72
QY 213 TAAGTCTCCAGATACAGACTGCAGGCAGACACCTCCTTCTCGTCTGCTGCTGACCTC 272
DB 73 TCAGCTCCGGAGATAGACTGCAAGCAGAACGCGCTTTCTGCTGCTGCTGACGCTC 132
QY 273 ATCCCAAAACAGTTGGCTGAGCGCATGCGCATCGGCAGAGCTGGGGGAAGAGAGGAC 332
DB 133 ATCCCAAGAGCTGGCGCTGCGATGCGCATCGGCAGAGCTGGGGTAGAGAGATC 192
QY 333 GGTGAAGGGAAGCAGCTGAAGACATTTCTTCTCTGGGGACCAACAGCAGTGCAGCGGA 392
DB 193 GGTGCAGGGCCAAACCGGTGAGGACCTTCTTCTCTGGGGTCTCAGACAGCAGCGGA 252
QY 393 AACAAAAGGTGGACAGAGGAGCAGCAGGACAGCGGACATATTCAGAGGATTCCT 452
DB 253 CATGACGCTACAGCCCTGGAGGAGGAGCAGCAGCCTGACATCATCCAGAGGATTCNA 312
QY 453 AGACGCTATTACATCTGACCTCAAGACCATGATGGCATAGATGGTCCATCGCTT 512
DB 313 GAGCCCTATTTCAACTTGACCTGAAGACCATGATGGGTATGGAATGGGTCTACACTT 372
QY 513 TTGTCCTCAGGGCGGTGTTGATGAACAAACAGACTCAGACATGTTTCATCAATGTTGACTA 572
DB 373 TTGTCCTCAGACCGCTTATGATGAACAAACGACTCAGACATGTTTGTGAATGTTGGCTA 432
QY 573 TCTGACTGAATGCT 587
DB 433 TCTGACCGAACTGCT 447

RESULT 10
A2768940/c
LOCUS A2768940 400 bp DNA linear GSS 16-FEB-2001
DEFINITION IM0569K09F Mouse 10kb plasmid UUC1M library Mus musculus genomic
clone UUC1M0569K09 F, genomic survey sequence.
ACCESSION A2768940
VERSION A2768940.1 GI:12888559
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 400)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
Niederhausern,A. and Wright,D.,Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0569 row: K column: 09
Seq primer: CGTTGTAACGACGCGCCAGT
Class: Plasmid ends
High quality sequence stop: 400.
Location/Qualifiers
1. 400
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUC1M0569K09"
/sex="Male"
/lab_hosts="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUC1M library"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
```





Sanger Institute  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: trop@anger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS SEQUENCE ID: ThDA01n08.plkSP6  
This sequence is from a Xenopus Gene Collection (XGC) library  
constructed by Nigel Garrett.  
Seq primer: SP6.

FEATURES  
Location/Qualifiers  
source  
1..629  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="THDA01n08"  
/dev stages="tailbud head (stage 28-30)"  
/lab host="Escherichia coli DH10B."  
/clone lib="XGC-tailbud-head"  
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dT primed from sug of poly A+ RNA from tailbud head. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN  
Query Match 18.2%; Score 184.2; DB 7; Length 629;  
Best Local Similarity 65.1%; Pred. No. 4.8e-43;  
Matches 287; Conservative 0; Mismatches 153; Indels 1; Gaps 1;  
QY 529 TTTGTGATGAAAAACAGACTCAGACATG-TTCATCAATGTTGACTATCTGACTGTAAGTCTCT 587  
DB 1 TTTGTGATGAGACGACTCAGACATGTTTGTTCACCTTCTACCTGGTCCAACTGCT 60  
QY 588 TCTGAAGAAAAACAGAACACACAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCC 647  
DB 61 GGCAAGAAAAACAGACTCTTCTAATTTTTTACTGGGTTTCTCAACTGAACGAGTACCC 120  
QY 648 CATAGGAGCCATTCAGCAAGTGTGTTGTCAGTAAATCTGAATATCCGTGGGACAGTA 707  
DB 121 GATAAGGAATATCTTCAGCAAGTGTGATCGCCAGTAAAGGGAATACCCAGGGGCCAAGTA 180  
QY 708 CCCACCATCTGCTCCGGCACCGGTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTA 767  
DB 181 CCCTCCATTTGTTGGGAGACTGATAGTCTTTCTGTAGACGTGCGCCAAAAGATCCA 240  
QY 768 CAATGTCTCCAGAGCGTCCCATACATTAATTAACATGGAAGACGTGTTTGTGGGGCTCTGCT 827  
DB 241 CAACATCTCCACGACAGTGCCTTTTTTCAACTGGAGAGCTCTATTTGGGGCTATGCT 300  
QY 828 CGAAGGCTGAACATCAGATTGGAGAGCTCCACTCCCGACCGACCTTTTTTTCAGGGGG 887  
DB 301 TGACATATTGACATTCATTGGAGGAACCTTCATACAGAGCAGACATTTCTTTGACAGAG 360  
QY 888 CTTAGGCTTCTCGGTATCGCTTCTCAGGAGGATCGTGGCCCTGCCACTTCATCAAGCCCTCG 947  
DB 361 CGAGTCATCTCCGTTTCAAAATACAGTAACTGTGACGTCCTCCCATGAGTCAAAACCATA 420  
QY 948 GACTCTCTGGACTACTGGCA 968  
DB 421 TGAGACATTGTATCTGAA 441

RESULT 13  
AY401366  
LOCUS  
DEFINITION Mus musculus B3GALT1 gene, VIRTUAL TRANSCRIPT, partial sequence,  
genomic survey sequence.  
ACCESSION AY401366  
VERSION AY401366.1 GI:39757355  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 981)

AUTHORS  
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,  
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,  
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,  
Adams, M.D. and Cargill, M.  
Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
Science 302 (5652), 1960-1963 (2003)  
14671302  
2 (bases 1 to 981)  
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,  
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,  
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,  
Adams, M.D. and Cargill, M.  
Direct Submission  
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
Location/Qualifiers  
source  
1..981  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/locus\_tag="B3GALT1"  
/gene="B3GALT1"  
/locus\_tag="HCM0866"

ORIGIN  
Query Match 18.0%; Score 182; DB 9; Length 981;  
Best Local Similarity 58.2%; Pred. No. 2.6e-42;  
Matches 361; Conservative 0; Mismatches 250; Indels 9; Gaps 2;  
QY 230 ACTCAGCAGCAGACACCTCCCTTCCTCGTCTGCTGGTGACCTCATCCCAACAAACAGTTGG 289  
DB 215 AGTCGGAGAAAAACATTCCTTTCTCGTGATCTCTATTAGCACACACACAAGGAATTG 274  
QY 290 CTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAAGAGAGGACGGTGAAGGGAAGACGAGC 349  
DB 275 ATGCCCGCCAGGCATCCGGGAGACATGGGGGATGAAAACAACATTCAAAGGGATCAAGA 334  
QY 350 TGAAGACATTTCTCTCTGGGGACCAACAGCAGTGCAGCGGAAACAAAGAGGTGGACC 409  
DB 335 TAGCCACACTTTCTCTGGGGCAAAATGCTGATCTCTGTAACAGATGGTGGAGC 394  
QY 410 AGGAGAGCCAGCAGACAGGGGACATATCCAGAAGGATTTCTTAGACGCTCTATTACAATC 469  
DB 395 AAGAGAGCCAGATCTTCATGACATCATCGTGGAGGACTTCATTGACTCTTACCACATC 454  
QY 470 TGACCCCTGAAGACCATGATGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGGGGCGT 529  
DB 455 TCACCCCTCAAAACCTTAATGGGATGAGATGGGTTCGCACCTTCTGTTCAAAAGGCCAAGT 514  
QY 530 TTGTGATGAAACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTC 589  
DB 515 ACGTGATGAAACCGACAGTGCACATTTTGTGAACATGGAACAACCTTATTATAAATCC 574  
QY 590 TGA-----GAAAAACAGAACCAACAGGTTTTTTCCTGGCTTCTTGAACCTCAATGAGT 643  
DB 575 TGAACCTCTACCAAGCCAGAGAGATTTTCACTGGTACGTCA---TCAAGCGG 631  
QY 644 TTCCCATCAGCAGCCATTTCAGCAAGTGGTTTGTTCAGTAAATCTGAATATCCGTGGGACA 703  
DB 632 GGCCAACTCAGGATGTCGGCAGTAAGTGGTATATGCGCTAGAGATTGTTACCTGCAGCA 691  
QY 704 GGTACCCACCATTTCTGCTCCGGCACCGGCTACGTGTTTTCTGGGACAGCTGGGAGTACAG 763  
DB 692 ACTACCCACCGTCTCTGTTTCAGGGGACTGGCTATATCTTTCCGCTGATGTGCTGAACATCA 751  
QY 764 TGTACAAATGCTCCCAAGAGCGTCCCATACATATTAACCTGGAAGAGCGTGTTCGTGGGGCTCT 823  
DB 752 TTTCAGAGACCTTCCTCCACAGCGGTGCTTTCATCTTGAAGATGTGTACGTGGGACTGT 811  
QY 824 GCCTCGAAGGCTGAACATC 843

```

Db      812  GTCTCGAAGCTGGCATC 831
|||||
|||
|||
|||
|||

RESULT 14
AJ507051
LOCUS      493 bp      mRNA      linear      EST 17-SEP-2002
DEFINITION Mus musculus BALB/c (Kahlem P) Mus musculus cDNA clone
182, mRNA sequence.
ACCESSION AJ507051
VERSION    AJ507051.1 GI:23034136
KEYWORDS   Mus musculus (house mouse)
SOURCE     Mus musculus
ORGANISM   Mus musculus

REFERENCE
AUTHORS    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 493)
Gitton,Y., Dahmane,N., Baik,S., Ruiz i Altaba,A., Neidhardt,L.,
Scholze,M., Hermann,B.G., Kahlem,P., Ben Kahla,A., Schrinner,S.,
Yildirimman,R., Herwig,R., Lehrach,H. and Yaspo,M.L.
A gene expression map of human chromosome 21 orthologs in the mouse
Nature (2002) In press
Contact: Kahlem P
Vertebrate Genomics
Max-Planck Institute for Molec. Genet.
Innestrasse 73, D-14195 Berlin, GERMANY.

FEATURES
source
1..493
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone="182"
/clone_lib="Mus musculus BALB/c (Kahlem P)"

ORIGIN
Query Match      17.1%; Score 173; DB 1; Length 493;
Best Local Similarity 73.9%; Pred. No. 9.6e-40;
Matches 232; Conservative 0; Mismatches 81; Indels 1; Gaps 1;

QY 671  GGTGTGTCAGTAAATCTGAATATCCGTGGGACAGGTA-CCACCAATCTGTCGGGACC 729
      |||||
Db 52  GGTGTGTCAGTAAAGTTCGAATATCCTCGGACAGGTACCCACCTTTTGTCTGTGTTACT 111

QY 730  GGCTACGTGTTTCTGGGCGAGTGGCGAGTCAGGTGTCATGTCCTCAAGAGCGTCCCA 789
      |||||
Db 112  GGTATGTGCTTTTCCAGGCGAGTGGCGGATCCAAGTATACAACGTCTCAGAGAGCGTTCCG 171

QY 790  TACATTAACTGGAGAGCGTGTGTTGCGGGCTCTGCCTCGAAAGGCTCAACATCAGATTG 849
      |||||
Db 172  TTCATCAAGCTGGAGGATGTGTTGTGGGCTCCGCTTGGCGAAGCTAAAGATCCGGCGC 231

QY 850  GAGGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCGCTC 909
      |||||
Db 232  GAGGAGCTGCACACCAACACAGACCTTCTCTCCCTGGCGGTTTACGCTTTCGTTGCGCGC 291

QY 910  TTCAGGAGATCGTGGCTCGCATTATCAAGCTCGGACTCTCTTGGACTACTGGCAG 969
      |||||
Db 292  TTTCANAAATTTGTGGCATGCCATTTTATGAAGCCCGGACCTGCTCACTTACTGGCAA 351

QY 970  GCTCTAGAGAAATTC 983
      |||||
Db 352  GCATCGGAGAACTC 365

RESULT 15
AK035215
LOCUS      2705 bp      mRNA      linear      HTC 03-APR-2004
DEFINITION Mus musculus adult male urinary bladder cDNA, RIKEN full-length
enriched library. clone:9530002C17 product:ODP-Gal:betaGlcNAc beta
1,3-galactosyltransferase, polypeptide 1, full insert sequence.
ACCESSION AK035215
VERSION    AK035215.1 GI:26084471

```

KEYWORDS  
SOURCE  
ORGANISM

HTC; CAP trapper.  
Mus musculus (house mouse)  
Mus musculus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

1 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636

2  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
11042159

3  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,  
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,  
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,  
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,  
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,  
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,  
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

4  
REFERENCE  
AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.

5  
REFERENCE  
AUTHORS

Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.

6  
REFERENCE  
AUTHORS

Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
(bases 1 to 2705)

Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,  
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,  
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,  
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,  
Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,  
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ono,M., Ohsato,N.,  
Okazaki,Y., Saito,K., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,  
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,  
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,  
Muramatsu,M. and Hayashizaki,Y.

Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,  
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL:http://genome.gsc.riken.jp/  
URL:http://fantom.gsc.riken.jp/.

FEATURES

Location/Qualifiers



This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2005, 14:50:12 ; Search time 2624.46 Seconds  
(without alignments)  
2575.655 Million cell updates/sec

Title: US-10-777-828-8

Perfect score: 1011

Sequence: 1 ccactcagcctctagcat.....aagattgcgcgtgtctga 1011

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 7400704 seqs, 3343079526 residues

Total number of hits satisfying chosen parameters: 14801408

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*\*

- 1: /cgm2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgm2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgm2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgm2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgm2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgm2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgm2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgm2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgm2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgm2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgm2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgm2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgm2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgm2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgm2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgm2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 17: /cgm2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgm2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*
- 19: /cgm2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq.\*
- 20: /cgm2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq.\*
- 21: /cgm2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq.\*
- 22: /cgm2\_6/ptodata/1/pubpna/US10J\_PUBCOMB.seq.\*
- 23: /cgm2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq.\*
- 24: /cgm2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 25: /cgm2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 26: /cgm2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1011	100.0	1011	US-10-777-828-8	Sequence 8, Appli
2	640	63.3	681	13 US-10-027-632-21208	Sequence 21208, A
3	640	63.3	681	13 US-10-027-632-21209	Sequence 21209, A
4	640	63.3	681	13 US-10-027-632-21210	Sequence 21210, A
5	640	63.3	681	17 US-10-027-632-21208	Sequence 21208, A
6	640	63.3	681	17 US-10-027-632-21209	Sequence 21209, A
7	640	63.3	681	17 US-10-027-632-21210	Sequence 21210, A

8	515.6	51.0	516	13	US-10-027-632-106210	Sequence 106210,
9	515.6	51.0	516	17	US-10-027-632-106210	Sequence 106210,
10	514	50.8	516	13	US-10-027-632-106208	Sequence 106208,
11	514	50.8	516	13	US-10-027-632-106209	Sequence 106209,
12	514	50.8	516	17	US-10-027-632-106208	Sequence 106208,
13	514	50.8	516	17	US-10-027-632-106209	Sequence 106209,
14	405.2	40.1	596	16	US-10-029-386-13121	Sequence 13121, A
15	161.8	16.0	2168	21	US-10-887-553A-603	Sequence 603, App
16	136.8	13.5	1965	21	US-10-764-420-1500	Sequence 1500, Ap
17	133.6	13.2	1358	20	US-10-370-715B-481	Sequence 481, App
18	133.6	13.2	1358	21	US-10-651-237-77	Sequence 77, Appl
19	133.6	13.2	1358	21	US-10-782-413-77	Sequence 77, Appl
20	133.6	13.2	1773	9	US-09-739-451-9	Sequence 9, Appli
21	133.6	13.2	2095	9	US-09-989-722-208	Sequence 208, App
22	133.6	13.2	2095	9	US-09-989-723-208	Sequence 208, App
23	133.6	13.2	2095	9	US-09-989-729-208	Sequence 208, App
24	133.6	13.2	2095	9	US-09-989-727-208	Sequence 208, App
25	133.6	13.2	2095	9	US-09-989-731-208	Sequence 208, App
26	133.6	13.2	2095	9	US-09-989-732-208	Sequence 208, App
27	133.6	13.2	2095	9	US-09-991-073-208	Sequence 208, App
28	133.6	13.2	2095	9	US-09-990-442-208	Sequence 208, App
29	133.6	13.2	2095	9	US-09-991-163-208	Sequence 208, App
30	133.6	13.2	2095	9	US-09-993-604-208	Sequence 208, App
31	133.6	13.2	2095	9	US-09-990-456-208	Sequence 208, App
32	133.6	13.2	2095	9	US-09-989-721-208	Sequence 208, App
33	133.6	13.2	2095	9	US-09-992-598-208	Sequence 208, App
34	133.6	13.2	2095	9	US-09-989-293A-208	Sequence 208, App
35	133.6	13.2	2095	9	US-09-989-735-208	Sequence 208, App
36	133.6	13.2	2095	9	US-09-990-444-208	Sequence 208, App
37	133.6	13.2	2095	9	US-09-991-181-208	Sequence 208, App
38	133.6	13.2	2095	9	US-09-989-730-208	Sequence 208, App
39	133.6	13.2	2095	9	US-09-990-436-208	Sequence 208, App
40	133.6	13.2	2095	9	US-09-993-687-208	Sequence 208, App
41	133.6	13.2	2095	10	US-09-989-734-208	Sequence 208, App
42	133.6	13.2	2095	10	US-09-997-653-208	Sequence 208, App
43	133.6	13.2	2095	10	US-09-989-724-208	Sequence 208, App
44	133.6	13.2	2095	10	US-09-989-728-208	Sequence 208, App
45	133.6	13.2	2095	10	US-09-990-441-208	Sequence 208, App

# ALIGNMENTS

## RESULT 1

US-10-777-828-8  
; Sequence 8, Application US/10777828  
; Publication No. US20040142425A1  
; GENERAL INFORMATION:  
; APPLICANT: Clausen, Henrik  
; APPLICANT: Amado, Margarita  
; TITLE OF INVENTION: UDP-GALACTOSE: BETA-N-ACETYL-GLUCOSAMINE BETA 1,3  
; TITLE OF INVENTION: GALACTOSYLTRANSFERASES, BETAGAL-T5  
; FILE REFERENCE: 7188-157  
; CURRENT APPLICATION NUMBER: US/10/777,828  
; PRIOR FILING DATE: 2004-02-12  
; PRIOR APPLICATION NUMBER: US/09/831,630  
; PRIOR FILING DATE: 2001-05-10  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 1011  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (79)..(1008)  
us-10-777-828-8

Query Match 100.0%; Score 1011; DB 19; Length 1011;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1011; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 CCACCTCAGCCTCCTAGCATAAACTAGACACATCCTCTTCGTTTGGAGTCTAATCATTT 60



Db 606 GACTACTGGCAGGCTCTAGAGAAATCCCGGGGGAAGATTGGCCGC 651  
 |||

RESULT 3

US-10-027-632-21209  
 ; Sequence 21209, Application US/10027632  
 ; Publication No. US2002019837A1

GENERAL INFORMATION

APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 ; TITLE OF INVENTION: Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 21209

LENGTH: 681

TYPE: DNA

ORGANISM: Human

US-10-027-632-21209

Query Match 63.3%; Score 640; DB 13; Length 681;  
 Best Local Similarity 99.1%; Pred. No. 1.2e-201;  
 Matches 640; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 358 TTCTTCTCTCTGGGACACACAGCAGTGCAGCGGAAACAAAGAGGTGGACAGGAGC 417  
 Db 6 TTCTTCTCTCTGGGACACACAGCAGTGCAGCGGAAACAAAGAGGTGGACAGGAGC 65  
 Qy 418 CAGGACACGGGGACATTTATCCAGAGGATTTCTAGACGCTATTACAATCTGACCTG 477  
 Db 66 CAGGACACGGGGACATTTATCCAGAGGATTTCTAGACGCTATTACAATCTGACCTG 125  
 Qy 478 AAGACCATGATGGGCATAGATGGGTCCATCGCTTTTCTCAGCGGGCGTTTGTGATG 537  
 Db 126 AAGACCATGATGGGCATAGATGGGTCCATCGCTTTTCTCAGCGGGCGTTTGTGATG 185  
 Qy 538 AAAACAGACTCAGACATGTTTCAATCAATGTGATCTGACTGACTGACTGACTGACTG 597  
 Db 186 AAAACAGACTCAGACATGTTTCAATCAATGTGATCTGACTGACTGACTGACTGACTG 245  
 Qy 598 AACAGAACACAGGTTTTCAGTGGCTTTTCAATCAATGTGATCTGACTGACTGACTGACTG 657  
 Db 246 AACAGAACACAGGTTTTCAGTGGCTTTTCAATCAATGTGATCTGACTGACTGACTGACTG 305  
 Qy 658 CCATTCAGCAAGTGGTTGTGATGAAATCTGAATATCCGTGGGACAGGTACCCACCATTC 717  
 Db 306 CCATTCAGCAAGTGGTTGTGATGAAATCTGAATATCCGTGGGACAGGTACCCACCATTC 365  
 Qy 718 TGCTCCGACACGGCTACGTTTCTGGGACAGTGGGAGTTCAGGTGATCAATGTCTTCC 777  
 Db 366 TGCTCCGACACGGCTACGTTTCTGGGACAGTGGGAGTTCAGGTGATCAATGTCTTCC 425  
 Qy 778 AAGAGCTCCCATACATTAATTAATCTGAAGACAGTGGTTTGTGGGCTCTGCTCGAAGGCTG 837  
 Db 426 AAGAGCTCCCATACATTAATTAATCTGAAGACAGTGGTTTGTGGGCTCTGCTCGAAGGCTG 485

Qy 838 AACATCAGATTGGAGGAGCTCCACTCCAGCGGACCTTTTTTCCAGGGGGCTTAGCGTTTC 897  
 Db 486 AACATCAGATTGGAGGAGCTCCACTCCAGCGGACCTTTTTTCCAGGGGGCTTAGCGTTTC 545  
 Qy 898 TCCGTATGCCCTCTTCAGGAGGATCGTGGCCCTGCCACTTCATCAAGCCTCGGACTCTCTTG 957  
 Db 546 TCCGTATGCCCTCTTCAGGAGGATCGTGGCCCTGCCACTTCATCAAGCCTCGGACTCTCTTG 605  
 Qy 958 GACTACTGGCAGGCTCTAGAGAAATCCCGGGGGAAGATTGTCCGC 1003  
 Db 606 GACTACTGGCAGGCTCTAGAGAAATCCCGGGGGAAGATTGTCCGC 651

RESULT 4

US-10-027-632-21210

; Sequence 21210, Application US/10027632

; Publication No. US2002019837A1

GENERAL INFORMATION

APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; TITLE OF INVENTION: Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 21210

LENGTH: 681

TYPE: DNA

ORGANISM: Human

US-10-027-632-21210

Query Match 63.3%; Score 640; DB 13; Length 681;  
 Best Local Similarity 99.1%; Pred. No. 1.2e-201;  
 Matches 640; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 358 TTCTTCTCTCTGGGACACACAGCAGTGCAGCGGAAACAAAGAGGTGGACAGGAGC 417  
 Db 6 TTCTTCTCTCTGGGACACACAGCAGTGCAGCGGAAACAAAGAGGTGGACAGGAGC 65  
 Qy 418 CAGGACACGGGGACATTTATCCAGAGGATTTCTAGACGCTATTACAATCTGACCTG 477  
 Db 66 CAGGACACGGGGACATTTATCCAGAGGATTTCTAGACGCTATTACAATCTGACCTG 125  
 Qy 478 AAGACCATGATGGGCATAGATGGGTCCATCGCTTTTCTCAGCGGGCGTTTGTGATG 537  
 Db 126 AAGACCATGATGGGCATAGATGGGTCCATCGCTTTTCTCAGCGGGCGTTTGTGATG 185  
 Qy 538 AAAACAGACTCAGACATGTTTCAATCAATGTGATCTGACTGACTGACTGACTGACTG 597  
 Db 186 AAAACAGACTCAGACATGTTTCAATCAATGTGATCTGACTGACTGACTGACTGACTG 245  
 Qy 598 AACAGAACACAGGTTTTCAGTGGCTTTTCAATCAATGTGATCTGACTGACTGACTGACTG 657  
 Db 246 AACAGAACACAGGTTTTCAGTGGCTTTTCAATCAATGTGATCTGACTGACTGACTGACTG 305  
 Qy 658 CCATTCAGCAAGTGGTTGTGATGAAATCTGAATATCCGTGGGACAGGTACCCACCATTC 717  
 Db 306 CCATTCAGCAAGTGGTTGTGATGAAATCTGAATATCCGTGGGACAGGTACCCACCATTC 365

QY 718 TGCTCCGACCGGCTACGTGTTTCTGGGACGTTGGGAGTCAAGTGTCAATGTCTCC 777  
 Db 366 TGCTCCGACCGGCTACGTGTTTCTGGGACGTTGGGAGTCAAGTGTCAATGTCTCC 425  
 QY 778 AAGAGCGTCCCATACATTAATACTGGAAGACGTGTTTGTGGGCTCTGCTCGAAGGCTG 837  
 Db 426 AAGAGCGTCCCATACATTAATACTGGAAGACGTGTTTGTGGGCTCTGCTCGAAGGCTG 485  
 QY 838 AACATCAGATTGGAGGAGCTCCATCCAGCCGACCTTTTTCAGGGGGCTTACGCTTC 897  
 Db 486 AACATCAGATTGGAGGAGCTCCATCCAGCCGACCTTTTTCAGGGGGCTTACGCTTC 545  
 QY 898 TCCGTATGCTCTTTCAGGAGGATCGTGCGCTGCCACTTTCATCAAGCCCTCGGACTCTCTTG 957  
 Db 546 TCCGTATGCTCTTTCAGGAGGATCGTGCGCTGCCACTTTCATCAAGCCCTCGGACTCTCTTG 605  
 QY 958 GACTACTGGCAGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCGCG 1003  
 Db 606 GACTACTGGCAGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCGCGC 651

RESULT 5

US-10-027-632-21208  
 ; Sequence 21208, Application US/10027632  
 ; Publication No. US20030204075A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 ; FILE REFERENCE: 108827.129  
 ; CURRENT APPLICATION NUMBER: US/10/027,632  
 ; CURRENT FILING DATE: 2002-04-30  
 ; PRIOR APPLICATION NUMBER: US 60/218,006  
 ; PRIOR FILING DATE: 2000-07-12  
 ; PRIOR APPLICATION NUMBER: US 60/198,676  
 ; PRIOR FILING DATE: 2000-04-20  
 ; PRIOR APPLICATION NUMBER: US 60/193,483  
 ; PRIOR FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: US 60/185,218  
 ; PRIOR FILING DATE: 2000-02-24  
 ; PRIOR APPLICATION NUMBER: US 60/167,363  
 ; PRIOR FILING DATE: 1999-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/156,358  
 ; PRIOR FILING DATE: 1999-08-09  
 ; NUMBER OF SEQ ID NOS: 325720  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 21208  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-10-027-632-21208

Query Match 63.3%; Score 640; DB 17; Length 681;  
 Best Local Similarity 99.1%; Pred. No. 1.2e-201;  
 Matches 640; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 358 TTCTTCCTCTCTGGGACCAACAGAGTGCAGCGGAAACAAAGAGGTGGACGAGAGC 417  
 Db 6 TTCTTCCTCTCTGGGACCAACAGAGTGCAGCGGAAACAAAGAGGTGGACGAGAGC 65  
 QY 418 CAGCGACACGGGACATTTATCCAGAGGATTTCTAGACGTCTATTCAATCTGACCCCTG 477  
 Db 66 CAGCGACACGGGACATTTATCCAGAGGATTTCTAGACGTCTATTCAATCTGACCCCTG 125  
 QY 478 AAGACCATGATGGGCATAGATGGGTCCATCGCTTTTGTCTCAGCGGCGCTTTGTGATG 537  
 Db 126 AAGACCATGATGGGCATAGATGGGTCCATCGCTTTTGTCTCAGCGGCGCTTTGTGATG 185  
 QY 538 AAAACAGACTCAGACATGTTTCATCAATGTGCACTATCTGACTGAACTGCTTCTGAAGAAA 597

Db 186 AAAACAGACTCAGACATGTTTTCATCAATGTGACTACTGACTGAACTGCTTCTGAAGAAA 245  
 QY 598 AACAGACACACGAGTGTTCCTCACTGGCTTCTTCAAACTCAATGAGTTTCCCATCAGGACG 657  
 Db 246 AACAGAACAAACGAGTGTTCCTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGACG 305  
 QY 658 CCATTTCAGCAAGTGGTGTTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTC 717  
 Db 306 CCATTTCAGCAAGTGGTGTTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTC 365  
 QY 718 TGCTCCGCAACCGGCTACGTTTTCCTGGGACGTCGTGGGAGTCAGGTGTCAATGTCTCC 777  
 Db 366 TGCTCCGCAACCGGCTACGTTTTCCTGGGACGTCGTGGGAGTCAGGTGTCAATGTCTCC 425  
 QY 778 AAGAGCGTCCCATACATTAACCTGGAAGACGTGTTTGTGGGCTCTGCTCGAAGGCTG 837  
 Db 426 AAGAGCGTCCCATACATTAACCTGGAAGACGTGTTTGTGGGCTCTGCTCGAAGGCTG 485  
 QY 838 AACATCAGATTGGAGGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTC 897  
 Db 486 AACATCAGATTGGAGGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTC 545  
 QY 898 TCCGTATGCTCTTTCAGGAGGATCGTGCGCTGCCACTTTCATCAAGCCCTCGGACTCTCTTG 957  
 Db 546 TCCGTATGCTCTTTCAGGAGGATCGTGCGCTGCCACTTTCATCAAGCCCTCGGACTCTCTTG 605  
 QY 958 GACTACTGGCAGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCGCG 1003  
 Db 606 GACTACTGGCAGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCGCGC 651

RESULT 6

US-10-027-632-21209  
 ; Sequence 21209, Application US/10027632  
 ; Publication No. US20030204075A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 ; FILE REFERENCE: 108827.129  
 ; CURRENT APPLICATION NUMBER: US/10/027,632  
 ; CURRENT FILING DATE: 2002-04-30  
 ; PRIOR APPLICATION NUMBER: US 60/218,006  
 ; PRIOR FILING DATE: 2000-07-12  
 ; PRIOR APPLICATION NUMBER: US 60/198,676  
 ; PRIOR FILING DATE: 2000-04-20  
 ; PRIOR APPLICATION NUMBER: US 60/193,483  
 ; PRIOR FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: US 60/185,218  
 ; PRIOR FILING DATE: 2000-02-24  
 ; PRIOR APPLICATION NUMBER: US 60/167,363  
 ; PRIOR FILING DATE: 1999-11-23  
 ; PRIOR APPLICATION NUMBER: US 60/156,358  
 ; PRIOR FILING DATE: 1999-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/146,002  
 ; PRIOR FILING DATE: 1999-08-09  
 ; NUMBER OF SEQ ID NOS: 325720  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 21209  
 ; LENGTH: 681  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-10-027-632-21209

Query Match 63.3%; Score 640; DB 17; Length 681;  
 Best Local Similarity 99.1%; Pred. No. 1.2e-201;  
 Matches 640; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 358 TTCTTCCTCTCTGGGACCAACAGAGTGCAGCGGAAACAAAGAGGTGGACGAGAGC 417  
 Db 6 TTCTTCCTCTCTGGGACCAACAGAGTGCAGCGGAAACAAAGAGGTGGACGAGAGC 65  
 QY 418 CAGCGACACGGGACATTTATCCAGAGGATTTCTAGACGTCTATTCAATCTGACCCCTG 477



```

Db 66 CAGGACACGGGGAGCATATCAGAAAGATTTCTAGACGCTATTCAATCTGACCCCTG 125
Qy 478 AAGACCATGATGGGCATAGATGGGTCCATCGCTTTTCTCTCAGGCGGGCGTTTGTGATG 537
Db 126 AAGACCATGATGGGCATAGATGGGTCCATCCTTTTCTCTCAGGCGGGCGTTTGTGATG 185
Qy 538 AAAACAGACTCAGACATGTTTCAATGTTGATCTGATCTGATGAACTCTCTGAAAGAA 597
Db 186 AAAACAGACTCAGACATGTTTCAATGTTGATCTGATCTGATGAACTCTCTGAAAGAA 245
Qy 598 AACAGAACACCGGTTTTCAGTGGCTTCTTGAACCTCAATGATTTCCCATCAGGCAG 657
Db 246 AACAGAACACCGGTTTTCAGTGGCTTCTTGAACCTCAATGATTTCCCATCAGGCAG 305
Qy 658 CCATTTCAGCAAGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTC 717
Db 306 CCATTTCAGCAAGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTC 365
Qy 718 TGCTCCGGCACCGGCTACGTTTCTGGGACGTTGTTTGTGGGGCTTGCCTCGAAAGGCTG 777
Db 366 TGCTCCGGCACCGGCTACGTTTCTGGGACGTTGTTTGTGGGGCTTGCCTCGAAAGGCTG 425
Qy 778 AAGAGCGTCCCATACATTAATCTGAAGACGTTGTTTGTGGGGCTTGCCTCGAAAGGCTG 837
Db 426 AAGAGCGTCCCATACATTAATCTGAAGACGTTGTTTGTGGGGCTTGCCTCGAAAGGCTG 485
Qy 838 AACATCAGATGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTC 897
Db 486 AACATCAGATGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTC 545
Qy 898 TCCGTATGCTCTTTTCAGGAGGATCGTGGCTGCCACTTTTCATCAAGCCCTCGGACTCTCTTG 957
Db 546 TCCGTATGCTCTTTTCAGGAGGATCGTGGCTGCCACTTTTCATCAAGCCCTCGGACTCTCTTG 605
Qy 958 GACTACTGGCAGGCTCTAGAGAAATTCCTGGGGGGAAGATTGTCCGC 1003
Db 606 GACTACTGGCAGGCTCTAGAGAAATTCCTGGGGGGAAGATTGTCCGC 651

```

```

RESULT 7
US-10-027-632-21210
; Sequence 21210, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21210
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-21210

```

```

Query Match 63.3%; Score 640; DB 17; Length 681;
Best Local Similarity 99.1%; Pred. No. 1.2e-201;
Matches 640; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 358 TTCTTCTCTCTGGGGACCCACGACGATGTCAGCGAAACAAAGAGGTGGACCGAGGAGC 417
Db 6 TTCTTCTCTCTGGGGACCCACGACGATGTCAGCGAAACAAAGAGGTGGACCGAGGAGC 65
Qy 418 CAGCGACACCGGGACATTTATCCAGAAGGATTTCTTAGACGCTATTATCAATCTGACCCCTG 477
Db 66 CAGCGACACCGGGACATTTATCCAGAAGGATTTCTTAGACGCTATTATCAATCTGACCCCTG 125
Qy 478 AAGACCATGATGGGCATAGATGGGTCCATCGCTTTTCTCTCAGGCGGGCGTTTGTGATG 537
Db 126 AAGACCATGATGGGCATAGATGGGTCCATCCTTTTCTCTCAGGCGGGCGTTTGTGATG 185
Qy 538 AAAACAGACTCAGACATGTTTCAATGTTGATCTGATCTGATGAACTCTCTGAAAGAA 597
Db 186 AAAACAGACTCAGACATGTTTCAATGTTGATCTGATCTGATGAACTCTCTGAAAGAA 245
Qy 598 AACAGAACACCGGTTTTCAGTGGCTTCTTGAACCTCAATGATTTCCCATCAGGCAG 657
Db 246 AACAGAACACCGGTTTTCAGTGGCTTCTTGAACCTCAATGATTTCCCATCAGGCAG 305
Qy 658 CCATTTCAGCAAGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTC 717
Db 306 CCATTTCAGCAAGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTC 365
Qy 718 TGCTCCGGCACCGGCTACGTTTCTGGGACGTTGTTTGTGGGGCTTGCCTCGAAAGGCTG 777
Db 366 TGCTCCGGCACCGGCTACGTTTCTGGGACGTTGTTTGTGGGGCTTGCCTCGAAAGGCTG 425
Qy 778 AAGAGCGTCCCATACATTAATCTGAAGACGTTGTTTGTGGGGCTTGCCTCGAAAGGCTG 837
Db 426 AAGAGCGTCCCATACATTAATCTGAAGACGTTGTTTGTGGGGCTTGCCTCGAAAGGCTG 485
Qy 838 AACATCAGATGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTC 897
Db 486 AACATCAGATGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTC 545
Qy 898 TCCGTATGCTCTTTTCAGGAGGATCGTGGCTGCCACTTTTCATCAAGCCCTCGGACTCTCTTG 957
Db 546 TCCGTATGCTCTTTTCAGGAGGATCGTGGCTGCCACTTTTCATCAAGCCCTCGGACTCTCTTG 605
Qy 958 GACTACTGGCAGGCTCTAGAGAAATTCCTGGGGGGAAGATTGTCCGC 1003
Db 606 GACTACTGGCAGGCTCTAGAGAAATTCCTGGGGGGAAGATTGTCCGC 651

```

```

RESULT 8
US-10-027-632-106210
; Sequence 106210, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002

```

```

; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 106210
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-106210

```

Query Match	51.0%;	Score 515.6;	DB 13;	Length 516;
Best Local Similarity	99.8%;	Pred. No. 2.6e-160;		
Matches 515;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	350	TGAAGACATTCTTCCTCTCGGGGACACACAGCAGTCGACGGGAACAAAAGAGCGTGGACC	409	
Db	1	TGAAGACATTCTTCCTCTCGGGGACACACAGCAGTCGACGGGAACAAAAGAGCGTGGACC	60	
Qy	410	AGGAGAGCCAGCGACACACGGGACATTATCCAGAAGAGGATTTCTTAGACGTCTATTACAATC	469	
Db	61	AGGAGAGCCAGCGACACGGGACATTATCCAGAAGAGGATTTCTTAGACGTCTATTACAATC	120	
Qy	470	TGACCCCTGAAGACCATGATGGGCATAGAATGGGTCATCGCTTTTGTCTCTCAGCGGGCGT	529	
Db	121	TGACCCCTGAAGACCATGATGGGCATAGAATGGGTCATCGCTTTTGTCTCTCAGCGGGCGT	180	
Qy	530	TTGTGATGAAAAACAGACTCAGACATGTTCAATCAATGTTGACTATCTCGACTGAACGTGCTC	589	
Db	181	TTGTGATGAAAAACAGACTCAGACATGTTCAATCAATGTTGACTATCTCGACTGAACGTGCTC	240	
Qy	590	TGAAGAAAAACAGAAACACACAGGTTTTTTCATCGGCTTCTTGAAACTCAAATGAGTGTCCCA	649	
Db	241	TGAAGAAAAACAGAAACACACAGGTTTTTTCATCGGCTTCTTGAAACTCAAATGAGTGTCCCA	300	
Qy	650	TCAGGCAGCCATTTCAGCAAGTGGTTGTCTAGTAAATCTGTAATATCCGTGGGACAGGTACC	709	
Db	301	TCAGGCAGCCATTTCAGCAAGTGGTTGTCTAGTAAATCTGTAATATCCGTGGGACAGGTACC	360	
Qy	710	CACCATCTCTCGGCACACGGCTACGTGTTTCTGGCGAGCTGGCGAGTCAGGTGTACA	769	
Db	361	CACCATCTCTCGGCACACGGCTACGTGTTTCTGGCGAGCTGGCGAGTCAGGTGTACA	420	
Qy	770	ATGTCTCCAAAGAGCGTCCCATACATTAAACTGGGAAGACGTGTTTGTGGGGCTCTGGCTCG	829	
Db	421	ATGTCTCCAAAGAGCGTCCCATACATTAAACTGGGAAGACGTGTTTGTGGGGCTCTGGCTCG	480	
Qy	830	AAAGGCTGAACATCAGATTGGAGAGCTCCACTCCC	865	
Db	481	AAAGGCTGAACATCAGATTGGAGAGCTCCACTCCC	516	

## RESULT 9

US-10-027-632-106210  
; Sequence 106210, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:

APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28

```

; PRIOR APPLICATION NUMBER: US 60/146,002
;
; PRIOR FILING DATE: 1999-08-09
;
; NUMBER OF SEQ ID NOS: 325720
;
; SOFTWARE: FastSeq for Windows Version 4.0
;
; SEQ ID NO 106210
;

```

Query Match	51.0%;	Score 515.6;	DB 17;	Length 516;
Best Local Similarity	99.8%;	Pred. No. 2.6e-160;		
Matches 515; Conservative	1;	Mismatches 0;	Indels 0;	

350	QY	TTGAAGACATTTCTTCTCTCTGGGGACCAACGACGAGTCAGCGGAAAACAAAGAGAGGTGGACC	409
1	DB	TTGAAGACATTTCTTCTCTCTGGGGACCAACGACGAGTCAGCGGAAAACAAAGAGAGGTGGACC	60
410	QY	AGGAGAGCCAGCGACACGGGGACATTTATCCAGAAGGATTTCTCTAGACGCTATTATCAATC	469
61	DB	AGGAGAGCCAGCGACACGGGGACATTTATCCAGAAGGATTTCTCTAGACGCTATTATCAATC	120
470	QY	TGACCCCTGAAGACCAATGATGGGCATAGAAATGGGTCCATCGCTTTTGTCTCTCAGGCGGCGT	529
121	DB	TGACCCCTGAAGACCAATGATGGGCATAGAAATGGGTCCATCGCTTTTGTCTCTCAGGCGGCGT	180
530	QY	TTGTGATCAAAAACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTC	589
181	DB	TTGTGATGAAAACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTC	240
590	QY	TGAAGAAAAACAGAAACACAGGTTTTTCACTGGCTCTTTGAAACTCAATGAGTTCCTCCA	649
241	DB	TGAAGAAAAACAGAAACACAGGTTTTTCACTGGCTCTTTGAAACTCAATGAGTTCCTCCA	300
650	QY	TCAGCAGCCCAATTCAGCAAGGTGGTTGTTCAGTAAATCTGAAATATCCGTGGGACAGGTACC	709
301	DB	TCAGCAGCCCAATTCAGCAAGGTGGTTGTTCAGTAAATCTGAAATATCCGTGGGACAGGTACC	360
710	QY	CACCAATTCGTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGGAGTCAGGTGTACA	769
361	DB	CACCAATTCGTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGGAGTCAGGTGTACA	420
770	QY	ATGCTCTCAAGAGCGTCCCATACATTAATTCGAAGACGTTGTTGTGGGGCTCTCGCTCG	829
421	DB	ATGCTCTCAAGAGCGTCCCATACATTAATTCGAAGACGTTGTTGTGGGGCTCTCGCTCG	480
830	QY	AAAGGCTGAAACATCAGATTTGGAGGAGCTCCACTCCC	865
481	DB	AAAGGCTGAAACATCAGATTTGGAGGAGCTCCACTCCC	516

## RESULT 10

US-10-027-632-106208  
; Sequence 106208, Application US/10027632  
; Publication No. US20020198371A1

```

1 / GENERAL INFORMATION:
2 / APPLICANT: Wang, David G.
3 / TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
4 / POLYMORPHISMS IN THE HUMAN GENOME
5 / FILE REFERENCE: 108827.129
6 / CURRENT APPLICATION NUMBER: US/10/027,632
7 / CURRENT FILING DATE: 2002-04-30
8 / PRIOR APPLICATION NUMBER: US 60/218,006
9 / PRIOR FILING DATE: 2000-07-12
10 / PRIOR APPLICATION NUMBER: US 60/198,676
11 / PRIOR FILING DATE: 2000-04-20
12 / PRIOR APPLICATION NUMBER: US 60/193,483
13 / PRIOR FILING DATE: 2000-03-29
14 / PRIOR APPLICATION NUMBER: US 60/185,218
15 / PRIOR FILING DATE: 2000-02-24
16 / PRIOR APPLICATION NUMBER: US 60/167,363
17 / PRIOR FILING DATE: 1999-11-23
18 / PRIOR APPLICATION NUMBER: US 60/156,358

```

```

RESULT 11
US-10-027-632-106209
; Sequence 106209, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23

```

; PRIOR FILING DATE: 1999-11-23  
 ; PRIOR APPLICATION NUMBER: US 60/156,358  
 ; PRIOR FILING DATE: 1999-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/146,002  
 ; PRIOR FILING DATE: 1999-08-09  
 ; PRIOR APPLICATION NUMBER: US 60/146,002  
 ; NUMBER OF SEQ ID NOS: 325720  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 106208  
 ; LENGTH: 516  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-10-027-632-106208

Query Match 50.8%; Score 514; DB 17; Length 516;  
 Best Local Similarity 99.6%; Pred. No. 9e-160;  
 Matches 514; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 350 TGAAGACATTCTTCTCTCTGGGACCAACAGCAGTCAGCGGAAACAAAGAGGTGGACC 409  
 Db 1 TGAAGACATTCTTCTCTCTGGGACCAACAGCAGTCAGCGGAAACAAAGAGGTGGACC 60

Qy 410 AGGAGAGCCAGCAGCAGCGGACATATCCAGAGGATTTCTAGACGTCTATTACAATC 469  
 Db 61 AGGAGAGCCAGCAGCAGCGGACATATCCAGAGGATTTCTAGACGTCTATTACAATC 120

Qy 470 TGACCCCTGAAGACCATGATGGGCATAGATGGGTCCATCGCTTTTGTCTCAGGGGGGT 529  
 Db 121 TGACCCCTGAAGACCATGATGGGCATAGATGGGTCCATCGCTTTTGTCTCAGGGGGGT 180

Qy 530 TTGTGATGAAACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTC 589  
 Db 181 TTGTGATGAAACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTC 240

Qy 590 TGAAGAAAAACAGAAACACAGGTTTTTCATCGGCTTCTTCAAACTCAATCAGTTTCCCA 649  
 Db 241 TGAAGAAAAACAGAAACACAGGTTTTTCATCGGCTTCTTCAAACTCAATCAGTTTCCCA 300

Qy 650 TCAGGAGCCATTCAGGAAGTGGTTGTGAGTAATCTGAATATCCGTGGGACAGGTACC 709  
 Db 301 TCAGGAGCCATTCAGGAAGTGGTTGTGAGTAATCTGAATATCCGTGGGACAGGTACC 360

Qy 710 CACCATTCGTCTCGGACCGGCTACGTGTTTCTGCGGACGTGGCGAGTCAGGTGTACA 769  
 Db 361 CACCATTCGTCTCGGACCGGCTACGTGTTTCTGCGGACGTGGCGAGTCAGGTGTACA 420

Qy 770 ATGTCTCAAGAGCGTCCCATACATTAACCTGGAAGACGTGTTTGTGGGGCTCTGCCTCG 829  
 Db 421 ATGTCTCAAGAGCGTCCCATACATTAACCTGGAAGACGTGTTTGTGGGGCTCTGCCTCG 480

Qy 830 AAAGCTGAACATCAGATTGGAGGAGCTCCACTCCC 865  
 Db 481 AAAGCTGAACATCAGATTGGAGGAGCTCCACTCCC 516

RESULT 13  
 US-10-027-632-106209  
 ; Sequence 106209, Application US/10027632  
 ; Publication No. US20030204075A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 ; TITLE OF INVENTION: Polymorphisms in the Human Genome  
 ; FILE REFERENCE: 108827.129  
 ; CURRENT APPLICATION NUMBER: US/10/027,632  
 ; CURRENT FILING DATE: 2002-04-30  
 ; PRIOR APPLICATION NUMBER: US 60/218,006  
 ; PRIOR FILING DATE: 2000-07-12  
 ; PRIOR APPLICATION NUMBER: US 60/198,676  
 ; PRIOR FILING DATE: 2000-04-20  
 ; PRIOR APPLICATION NUMBER: US 60/193,483  
 ; PRIOR FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: US 60/185,218  
 ; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363  
 ; PRIOR FILING DATE: 1999-11-23  
 ; PRIOR APPLICATION NUMBER: US 60/156,358  
 ; PRIOR FILING DATE: 1999-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/146,002  
 ; PRIOR FILING DATE: 1999-08-09  
 ; NUMBER OF SEQ ID NOS: 325720  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 106209  
 ; LENGTH: 516  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-10-027-632-106209

Query Match 50.8%; Score 514; DB 17; Length 516;  
 Best Local Similarity 99.6%; Pred. No. 9e-160;  
 Matches 514; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 350 TGAAGACATTCTTCTCTCTGGGACCAACAGCAGTCAGCGGAAACAAAGAGGTGGACC 409  
 Db 1 TGAAGACATTCTTCTCTCTGGGACCAACAGCAGTCAGCGGAAACAAAGAGGTGGACC 60

Qy 410 AGGAGAGCCAGCAGCAGCGGACATATCCAGAGGATTTCTAGACGTCTATTACAATC 469  
 Db 61 AGGAGAGCCAGCAGCAGCGGACATATCCAGAGGATTTCTAGACGTCTATTACAATC 120

Qy 470 TGACCCCTGAAGACCATGATGGGCATAGATGGGTCCATCGCTTTTGTCTCAGGGGGGT 529  
 Db 121 TGACCCCTGAAGACCATGATGGGCATAGATGGGTCCATCGCTTTTGTCTCAGGGGGGT 180

Qy 530 TTGTGATGAAACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTC 589  
 Db 181 TTGTGATGAAACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTC 240

Qy 590 TGAAGAAAAACAGAAACACAGGTTTTTCATCGGCTTCTTCAAACTCAATCAGTTTCCCA 649  
 Db 241 TGAAGAAAAACAGAAACACAGGTTTTTCATCGGCTTCTTCAAACTCAATCAGTTTCCCA 300

Qy 650 TCAGGAGCCATTCAGGAAGTGGTTGTGAGTAATCTGAATATCCGTGGGACAGGTACC 709  
 Db 301 TCAGGAGCCATTCAGGAAGTGGTTGTGAGTAATCTGAATATCCGTGGGACAGGTACC 360

Qy 710 CACCATTCGTCTCGGACCGGCTACGTGTTTCTGCGGACGTGGCGAGTCAGGTGTACA 769  
 Db 361 CACCATTCGTCTCGGACCGGCTACGTGTTTCTGCGGACGTGGCGAGTCAGGTGTACA 420

Qy 770 ATGTCTCAAGAGCGTCCCATACATTAACCTGGAAGACGTGTTTGTGGGGCTCTGCCTCG 829  
 Db 421 ATGTCTCAAGAGCGTCCCATACATTAACCTGGAAGACGTGTTTGTGGGGCTCTGCCTCG 480

Qy 830 AAAGCTGAACATCAGATTGGAGGAGCTCCACTCCC 865  
 Db 481 AAAGCTGAACATCAGATTGGAGGAGCTCCACTCCC 516

RESULT 14  
 US-10-029-386-13121  
 ; Sequence 13121, Application US/10029386  
 ; Publication No. US20030194704A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharron G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI  
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
 ; FILE REFERENCE: AROMICA-X-2  
 ; CURRENT APPLICATION NUMBER: US/10/029,386  
 ; CURRENT FILING DATE: 2001-12-20  
 ; NUMBER OF SEQ ID NOS: 34288  
 ; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 13121  
 ; LENGTH: 596  
 ; TYPE: DNA



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2005, 05:52:47 ; Search time 603.053 Seconds  
(without alignments)  
9924.272 Million cell updates/sec

Title: US-10-777-828-8

Perfect score: 1011

Sequence: 1 ccacctcagcctctagcat.....aagattgcgcctgtctga 1011

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04.\*

1: Geneseqn1980s.\*

2: Geneseqn1990s.\*

3: Geneseqn2000s.\*

4: Geneseqn2001as.\*

5: Geneseqn2001bs.\*

6: Geneseqn2002as.\*

7: Geneseqn2002bs.\*

8: Geneseqn2003as.\*

9: Geneseqn2003bs.\*

10: Geneseqn2003cs.\*

11: Geneseqn2003ds.\*

12: Geneseqn2004as.\*

13: Geneseqn2004bs.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1011	100.0	1011	3	AAA27959 Human bet
2	1006.2	99.5	3409	10	ABZ56919 B3GALT nu
3	1006.2	99.5	10562	3	AAA93876 Human bet
4	928.8	91.9	933	10	ABZ56916 B3GALT nu
5	928.6	91.8	2629	10	ABZ56918 B3GALT nu
6	928.6	91.8	2762	10	ABZ56917 B3GALT nu
7	928.6	91.8	2775	3	AAA93875 Human bet
8	928.2	91.8	933	12	ADK68456 Human bet
9	405.2	40.1	596	12	ACH79926 Human gen
10	163.4	16.2	1037	10	ABZ56909 B3GALT nu
11	163.4	16.2	1739	2	AAQ67067 Beta-1,3-
12	161.8	16.0	2168	10	ABZ56910 B3GALT nu
13	133.6	13.2	1092	12	ADO00390 Novel hum
14	133.6	13.2	1092	12	ADN98821 Novel hum
15	133.6	13.2	1358	10	ADF76806 Novel hum
16	133.6	13.2	1773	2	AAx87193 Human Den
17	133.6	13.2	1897	6	ABK51201 Human CDN
18	133.6	13.2	1897	12	ADO19801 Human PRO
19	133.6	13.2	2095	3	Az65022 Membrane-
20	133.6	13.2	2095	4	AAs46005 Human DNA

#### ALIGNMENTS

##### RESULT 1

AAA27959

ID AAA27959 standard; DNA; 1011 BP.

XX

AC AAA27959;

XX

DT 15-AUG-2000 (first entry)

XX

DE Human beta3Gal-T5 gene sequence.

XX

KW UDP-D-galactose; beta-N-acetylglucosamine beta1,3-galactosyltransferase;

KW Beta3gal-T5; ss; human; chromosome 21q22.3; galactosylation;

KW beta1,3-galactosyl glycosylated saccharide production; glycopeptide;

KW glycoprotein.

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 79..1011

FT /\*tag= a

FT /\*product= "Beta3Gal-T5"

FT complement(79..98)

FT /\*tag= b

FT primer\_bind complement(150..170)

FT /\*tag= c

FT primer\_bind 991..1011

FT /\*tag= d

PN WO200029558-A1.

XX

PD 25-MAY-2000.

XX

PF 11-NOV-1999; 99WO-US026807.

XX

PR 13-NOV-1998; 98DK-00001483.

XX

PA (CLAU/) CLAUSEN H.

XX

PI Clausen H, Amado M;

XX

DR WPI; 2000-399728/34.

DR P-PSDB; AAY94641.

XX

PT Novel nucleic acid sequence encoding human UDP-galactose:beta-N-  
PT acetylglucosamine beta1,3-galactosyltransferase useful for obtaining beta  
PT 1,3-galactosyl glycosylated saccharides and glycopeptides or  
PT glycoproteins.

PS Claim 7; Fig 1; 74pp; English.

XX  
XX The present invention relates to a nucleic acid sequence encoding UDP-D-  
CC galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase  
CC (beta3Gal-T5). Beta3 transferases add galactose to the hydroxy group at  
CC carbon 3 of 2-acetamido-2-deoxy-D-glucose (GlcNAc). The present sequence  
CC represents the human beta3Gal-T5 gene sequence. The beta3Gal-T5 gene is  
CC located on human chromosome 21q22.3. Beta3Gal-T5 is a type II  
CC transmembrane glycoprotein. The invention also relates to the beta3Gal-T5  
CC protein sequence, a nucleic acid vector comprising the beta3Gal-T5  
CC nucleotide sequence, a host cell comprising the vector, and a method for  
CC the production of the beta3Gal-T5 protein from the host cells. The  
CC methods of the invention can be used for recombinant production of  
CC beta3Gal-T5 for use as a catalyst and for recombinant production of  
CC peptides or proteins with appropriate galactosylation. The beta3Gal-T5  
CC protein can be used to obtain beta1,3-galactosyl glycosylated  
CC saccharides, glycopeptides or glycoproteins  
XX  
SQ Sequence 1011 BP; 247 A; 256 C; 251 G; 257 T; 0 U; 0 Other;

Query Match 100.0%; Score 1011; DB 3; Length 1011;  
Best Local Similarity 100.0%; Pred. No. 5e-304;  
Matches 1011; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCTCAGCCTCCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGTCTAATCATTT 60  
DB 1 CCACCTCAGCCTCCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGTCTAATCATTT 60  
QY 61 GGATTTTGTCTTTCAGATGGCTTCCGAGATGAGATTGATATATCTGCTTTCTG 120  
DB 61 GGATTTTGTCTTTCAGATGGCTTCCGAGATGAGATTGATATATCTGCTTTCTG 120  
QY 121 GTTCTGGGGCTCTTTGTTTATTTTATGATGATACAGTCTAAATCTTTCAAAGAACAG 180  
DB 121 GTTCTGGGGCTCTTTGTTTATTTTATGATGATACAGTCTAAATCTTTCAAAGAACAG 180  
QY 181 TCCTTTGTTTACAAGAACAGCGGAATCTTCTTAAGCTCCAGATACAGACTGCAGGCAG 240  
DB 181 TCCTTTGTTTACAAGAACAGCGGAATCTTCTTAAGCTCCAGATACAGACTGCAGGCAG 240  
QY 241 ACACCTCCTCTTCTCGTCTGCTGCTGACCTCATCCCAAAACAGATTGGCTGAGCGCATG 300  
DB 241 ACACCTCCTCTTCTCGTCTGCTGCTGACCTCATCCCAAAACAGATTGGCTGAGCGCATG 300  
QY 301 GCCATCCGGCAGACGTGGGGGAAAGAGAGAGACGGTGAAGGGAAAGCAGCTGAAGCATTC 360  
DB 301 GCCATCCGGCAGACGTGGGGGAAAGAGAGAGACGGTGAAGGGAAAGCAGCTGAAGCATTC 360  
QY 361 TTCTCTCTGGGGACCCACAGCAGTGCAGCGAAACAAAGAGGTGGACAGAGAGCCAG 420  
DB 361 TTCTCTCTGGGGACCCACAGCAGTGCAGCGAAACAAAGAGGTGGACAGAGAGCCAG 420  
QY 421 CGACACGGGGACATTATTCAGAGAGATTTCCTAGACGCTTATTAACTCTGACCTGAAG 480  
DB 421 CGACACGGGGACATTATTCAGAGAGATTTCCTAGACGCTTATTAACTCTGACCTGAAG 480  
QY 481 ACCATGATGGGCATAGATGGTTCATCGCTTTTGTCTCAGGGCGGCTTTGTGATGAA 540  
DB 481 ACCATGATGGGCATAGATGGTTCATCGCTTTTGTCTCAGGGCGGCTTTGTGATGAA 540  
QY 541 ACAGACTCAGACATCTTTCATCAATCTTGACTATCTGACTGAACTGCTTCTGAAGAAAAC 600  
DB 541 ACAGACTCAGACATCTTTCATCAATCTTGACTATCTGACTGAACTGCTTCTGAAGAAAAC 600  
QY 601 AGAACACACAGGTTTTTCACTGGCTTTCTTGAACTCAATGATGTTTCCCATCAGGCAGCCA 660  
DB 601 AGAACACACAGGTTTTTCACTGGCTTTCTTGAACTCAATGATGTTTCCCATCAGGCAGCCA 660

QY 661 TTCAGCAAGTGGTTTGTCAATAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGC 720  
DB 661 TTCAGCAAGTGGTTTGTCAATAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGC 720  
QY 721 TCCGGCACCCGGCTACGTTGTTTCTTGGCGACGTCGGCGAGTCAGGTGTACAATGTCTCCAAG 780  
DB 721 TCCGGCACCCGGCTACGTTGTTTCTTGGCGACGTCGGCGAGTCAGGTGTACAATGTCTCCAAG 780  
QY 781 AGCGTCCCATACATTAACCTGGAAGACGTTGTTGTTGGGGCTCTGCCTCGAAAGCTGAAC 840  
DB 781 AGCGTCCCATACATTAACCTGGAAGACGTTGTTGTTGGGGCTCTGCCTCGAAAGCTGAAC 840  
QY 841 ATCAGATTGGAGGAGCTCCCACTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTCTCC 900  
DB 841 ATCAGATTGGAGGAGCTCCCACTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTCTCC 900  
QY 901 GTATGCTCTTTCAGGAGGATCGTGGCTGCGACCTTCATCAAGCCCTCGGACTCTCTTGGAC 960  
DB 901 GTATGCTCTTTCAGGAGGATCGTGGCTGCGACCTTCATCAAGCCCTCGGACTCTCTTGGAC 960  
QY 961 TACTGGCAGGCTCTAGAGAAATCCCGGGGGGAGATTGTCCGCTGTCTGA 1011  
DB 961 TACTGGCAGGCTCTAGAGAAATCCCGGGGGGAGATTGTCCGCTGTCTGA 1011

RESULT 2  
ABZ56919  
ID ABZ56919 standard; DNA; 3409 BP.

XX ABZ56919;  
XX AC  
XX AC  
DT 04-APR-2003 (first entry)  
XX  
DE B3GALT nucleic acid sequence # SEQ ID 11.  
XX B3GALT; beta-1,3 galactosyltransferase; p53; cancer; breast; colon;  
KW kidney; lung; ovary; gene; ds.  
XX Homo sapiens.

OS  
FN WO200299044-A2.  
PN  
PD 12-DEC-2002.

XX  
PP 02-JUN-2002; 2002WO-US017356.  
XX  
PR 05-JUN-2001; 2001US-0296076P.  
PR 10-OCT-2001; 2001US-0328605P.  
PR 15-FEB-2002; 2002US-0357253P.

XX  
XX (EXEL-) EXELIXIS INC.  
FA  
XX  
PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
PI WPI; 2003-156849/15.

XX  
DR  
XX Identifying p53 pathway modulating agents with B3GALT genes, useful for  
PT the diagnosis and treatment of disorders associated with defects in the  
PT p53 pathway, such as cancer of the breast, colon, kidneys, lung and  
PT ovary.  
XX  
XX Disclosure; Page 55-57; 82pp; English.

PS  
XX The invention relates to identifying a candidate p53 pathway modulating  
CC agent in humans that is referred to in the specification as B3GALT (beta-  
CC 1,3 galactosyltransferase). Also disclosed is a method for diagnosing a  
CC disease in a patient, by contacting a sample with a probe for B3GALT  
CC expression, and comparing the results with a control, and determining  
CC whether the results indicate a likelihood of disease. Methods and  
CC compositions of the invention are useful for the diagnosis and treatment  
CC of disorders associated with defects in the p53 pathway, such as cancer  
CC of the breast, colon, kidneys, lung and ovary. The current sequence  
CC represents a B3GALT nucleic acid sequence referred to in the disclosure



CC of the invention  
SQ Sequence 3409 BP; 833 A; 842 C; 833 G; 901 T; 0 U; 0 Other;  
Query Match 99.5%; Score 1006.2; DB 10; Length 3409;  
Best Local Similarity 99.7%; Pred. No. 3.2e-302;  
Matches 1008; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CCACCTCAGCCTCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGCTTAATCATTT 60  
Db 1003 CCACCTCAGCCTCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGCTTAATCATTT 1062  
QY 61 GGATTTTGTCTTTTCAGATGCTTTCCGAAGATGAGATTGATATATCTGCCCTTCG 120  
Db 1063 GGATTTTGTCTTTTCAGATGCTTTCCGAAGATGAGATTGATATATTTGCCCTTCG 1122  
QY 121 GTTCTGGGGCTCTTTGTTGTTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAG 180  
Db 1123 GTTCTGGGGCTCTTTGTTGTTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAG 1182  
QY 181 TCCTTTGTTTACAAGAAAGACGGGAATCTTCTTAAGCTCCAGATACAGACTGCAGGCAG 240  
Db 1183 TCCTTTGTTTACAAGAAAGACGGGAATCTTCTTAAGCTCCAGATACAGACTGCAGGCAG 1242  
QY 241 ACACCTCCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
Db 1243 ACACCTCCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1302  
QY 301 GCCATCCGCGACAGCTGGGGGAAAGAGAGAGCGGTGAAGGGAAGACAGCTGAAGACATTC 360  
Db 1303 GCCATCCGCGACAGCTGGGGGAAAGAGAGAGCGGTGAAGGGAAGACAGCTGAAGACATTC 1362  
QY 361 TTCTCTCTGGGACACACAGCAGTGCAGCGGAAACAAAAGAGGTGGACACGAGAGCCAG 420  
Db 1363 TTCTCTCTGGGACACACAGCAGTGCAGCGGAAACAAAAGAGGTGGACACGAGAGCCAG 1422  
QY 421 CGACACGGGGACATTTATCAGAAGGATTTCTAGACGCTATTACAACTCTGACCCCTGAAG 480  
Db 1423 CGACACGGGGACATTTATCAGAAGGATTTCTAGACGCTATTACAACTCTGACCCCTGAAG 1482  
QY 481 ACCATGATGGCAGATAGATGGTTCATCGCTTTGCTCTCAGGGGGCTTTGTGATGAAA 540  
Db 1483 ACCATGATGGCAGATAGATGGTTCATCGCTTTGCTCTCAGGGGGCTTTGTGATGAAA 1542  
QY 541 ACAGACTCAGACATGTTTCATCAATCTGCTATCTGACTGTAAGTCTTCTGAAGAAAC 600  
Db 1543 ACAGACTCAGACATGTTTCATCAATCTGCTATCTGACTGTAAGTCTTCTGAAGAAAC 1602  
QY 601 AGAACACACAGGTTTTCTCAGTGGCTTTCTGAAACTCAATGAGTTTCCCATCAGGCAGCA 660  
Db 1603 AGAACACACAGGTTTTCTCAGTGGCTTTCTGAAACTCAATGAGTTTCCCATCAGGCAGCA 1662  
QY 661 TTCAGCAGTGGTTGTCAGTAATCTGAATATCTCGTGGGACAGTACCCACCATTCGTC 720  
Db 1663 TTCAGCAGTGGTTGTCAGTAATCTGAATATCTCGTGGGACAGTACCCACCATTCGTC 1722  
QY 721 TCCGGCACCGGTACGCTGTTTCTGGCAGCTGGGAGTCAAGTGTCAATGTTCTCCAAG 780  
Db 1723 TCCGGCACCGGTACGCTGTTTCTGGCAGCTGGGAGTCAAGTGTCAATGTTCTCCAAG 1782  
QY 781 AGCGTCCCATACATTAATCTGAAGACGTTGTTGTTGGGCTCTGCTCCGAAGGCTGAAC 840  
Db 1783 AGCGTCCCATACATTAATCTGAAGACGTTGTTGTTGGGCTCTGCTCCGAAGGCTGAAC 1842  
QY 841 ATCAGATTGAGGAGCTCCATCCACGCGGACCTTTTTCAGGGGGCTTACGCTTCTCC 900  
Db 1843 ATCAGATTGAGGAGCTCCATCCACGCGGACCTTTTTCAGGGGGCTTACGCTTCTCC 1902  
QY 901 GTATGCTCTTCAGGAGATCGTGCCCTGCCATCTCATCAAGCCTCGACTCTCTTGAC 960  
Db 1903 GTATGCTCTTCAGGAGATCGTGCCCTGCCATCTCATCAAGCCTCGACTCTCTTGAC 1962  
QY 961 TACTGGCAGGCTCTAGAGAAATTCGCGGGGGAAGATTGTCCGCTGTCTGA 1011

Db 1963 TACTGGCAGGCTCTAGAGAAATTCGCGGGGGAAGATTGTCCGCTGTCTGA 2013  
RESULT 3  
AAA93876  
ID AAA93876 standard; DNA; 10562 BP.  
XX AAA93876;  
AC AC  
DT 15-JAN-2001 (first entry)  
XX Human beta3Gal-T5 encoding DNA.  
DE Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;  
KW digestive system; beta3Gal-T5; ds.  
KW Homo sapiens.  
OS  
XX WO200050608-A1.  
PN 31-AUG-2000.  
XX 24-FEB-2000; 2000WO-JP001070.  
PF 25-FEB-1999; 99JP-00047571.  
PR (KYOW ) KYOWA HAKKO KOGYO KK.  
PA Narimatsu H, Ieshiki S, Togayachi A, Sasaki K;  
PI WPI; 2000-549409/50.  
XX Beta-1,3 galactose transferase and DNA encoding it, useful for synthesis  
of type 1 sialyl Lewis, a carbohydrate for treatment of digestive system  
cancer.  
Claim 31; Page 103-111; 123pp; Japanese.  
This invention relates to a polypeptide (I) with beta-1,3 galactose  
transferase activity, or variants of (I) comprising amino acid additions,  
deletions and/or substitutions. Included in the invention is DNA encoding  
all or part of (I); expression vectors containing the DNA, host cells  
transformed by the vectors; a method for the preparation of the  
polypeptide by culture of the transformants or by expression in the milk  
of a transgenic mammal, and antibodies recognising (I). The Beta-1,3  
galactose transferase protein transfers galactose by beta-1,3 bonding to  
N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as  
GlcNAc-beta1-3Gal-beta1-4Glc) to give Gal-beta1-3GlcNAc. The protein and DNA  
encoding it are useful for the treatment and diagnosis of cancer of the  
digestive system. The present sequence represents a Beta3Gal-T5 encoding  
DNA sequence  
SQ Sequence 10562 BP; 2610 A; 2415 C; 2574 G; 2963 T; 0 U; 0 Other;  
Query Match 99.5%; Score 1006.2; DB 3; Length 10562;  
Best Local Similarity 99.7%; Pred. No. 6.3e-302;  
Matches 1008; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CCACCTCAGCCTCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGCTTAATCATTT 60  
Db 8156 CCACCTCAGCCTCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGCTTAATCATTT 8215  
QY 61 GGATTTTGTCTTTTCAGATGCTTTCCGAAGATGAGATTGATATATCTGCCCTTCG 120  
Db 8216 GGATTTTGTCTTTTCAGATGCTTTCCGAAGATGAGATTGATATATCTGCCCTTCG 8275  
QY 121 GTTCTGGGGCTCTTTGTTGTTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAG 180  
Db 8276 GTTCTGGGGCTCTTTGTTGTTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAG 8335  
QY 181 TCCTTTGTTTACAAGAAAGACGGGAATCTTCTTAAGCTCCAGATACAGACTGCAGGCAG 240

Db 8336 TCCTTTGTTTACAAGAAAGACGGGAATCTTCTTAAGCTCCAGATACAGACTGCAGGCAG 8395  
 Qy 241 ACACCTCCCTTCTCGTCTCTGCTGTGTAACCTCATCCCAAAACAGTTGGCTGAGCGCATG 300  
 Db 8396 ACACCTCCCTTCTCGTCTCTGCTGTGTAACCTCATCCCAAAACAGTTGGCTGAGCGCATG 8455  
 Qy 301 GCCATCCGGCAGACGTGGGGGAAAGAGAGGACGGTGAAGGGGAAAGCAGCTGAAGACATTC 360  
 Db 8456 GCCATCCGGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGGAAAGCAGCTGAAGACATTC 8515  
 Qy 361 TTCTCTCTGGGGACACACAGCAGTGCAGCGGAAACAAAAGAGGTGGACAGGAGAGCCAG 420  
 Db 8516 TTCTCTCTGGGGACACACAGCAGTGCAGCGGAAAGAAAGAGGTGGACAGGAGAGCCAG 8575  
 Qy 421 CGACACGGGACATTAATCCAGAAGATTTCCTAGACGTCTATTACAACTCTGACCCCTGAAG 480  
 Db 8576 CGACACGGGACATTAATCCAGAAGATTTCCTAGACGTCTATTACAACTCTGACCCCTGAAG 8635  
 Qy 481 ACCATGATGGGCATAGATGGTCCATCGCTTTGTCCTCAGGGCGGCTTTGTGATGAAA 540  
 Db 8636 ACCATGATGGGCATAGATGGTCCATCGCTTTGTCCTCAGGGCGGCTTTGTGATGAAA 8695  
 Qy 541 ACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAAAAC 600  
 Db 8696 ACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAAAAC 8755  
 Qy 601 AGAACAAACAGGTTTTTCACTGGCTTCTTGAAGAACTCAATGATGTTTCCCATCAGGCGCCA 660  
 Db 8756 AGAACAAACAGGTTTTTCACTGGCTTCTTGAAGAACTCAATGATGTTTCCCATCAGGCGCCA 8815  
 Qy 661 TTCAGCAGTGGTTTGTTCAGTAATCTCAATATCCGTGGGACAGTACCCACCATTCCTGC 720  
 Db 8816 TTCAGCAGTGGTTTGTTCAGTAATCTCAATATCCGTGGGACAGTACCCACCATTCCTGC 8875  
 Qy 721 TCCGCGACCGGCTACGTTGTTTCTGGCGACGTGGCGAGTCAAGTGTCAATGTTCTCCAAG 780  
 Db 8876 TCCGCGACCGGCTACGTTGTTTCTGGCGACGTGGCGAGTCAAGTGTCAATGTTCTCCAAG 8935  
 Qy 781 AGCGTCCCATACATTAACCTGGAAGACGTGTTTGTGGGCTCTGCTCGAAGGCTGAAC 840  
 Db 8936 AGCGTCCCATACATTAACCTGGAAGACGTGTTTGTGGGCTCTGCTCGAAGGCTGAAC 8995  
 Qy 841 ATCAGATTGGAGGAGCTCCATCCCGACCGACCTTTTTCAGGGGGCTTACGCTTCTCC 900  
 Db 8996 ATCAGATTGGAGGAGCTCCATCCCGACCGACCTTTTTCAGGGGGCTTACGCTTCTCC 9055  
 Qy 901 GTATGCTCTTTCAGGAGGATCGTGGCTGCGACATTCATCAAGCCCTCGGACTCTCTTGGAC 960  
 Db 9056 GTATGCTCTTTCAGGAGGATCGTGGCTGCGACATTCATCAAGCCCTCGGACTCTCTTGGAC 9115  
 Qy 961 TACTGGCAGGCTCTAGAGAAATTCGCGGGGGAAGATTGCGCCCTGTCGA 1011  
 Db 9116 TACTGGCAGGCTCTAGAGAAATTCGCGGGGGAAGATTGCGCCCTGTCGA 9166

RESULT 4  
 ABZ56916  
 ID ABZ56916 standard; DNA; 933 BP.

XX AC ABZ56916;  
 XX  
 DT 04-APR-2003 (first entry)  
 XX B3GALT nucleic acid sequence # SEQ ID 8.  
 DE  
 DE B3GALT; beta-1,3 galactosyltransferase; p53; cancer; breast; colon;  
 KW kidney; lung; ovary; gene; ds.  
 KW  
 KW  
 OS Homo sapiens.  
 XX  
 XX WO200299044-A2.  
 XX  
 XX 12-DEC-2002.

XX 02-JUN-2002; 2002WO-US017356.  
 XX 05-JUN-2001; 2001US-0296076P.  
 PR 10-OCT-2001; 2001US-0328605P.  
 PR 15-FEB-2002; 2002US-0357253P.  
 XX (EXEL-) EXELIXIS INC.  
 XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
 FI WPI; 2003-156849/15.  
 XX  
 DR Identifying p53 pathway modulating agents with B3GALT genes, useful for  
 XX the diagnosis and treatment of disorders associated with defects in the  
 PT p53 pathway, such as cancer of the breast, colon, kidneys, lung and  
 PT ovary.  
 PT  
 XX Disclosure; Page 51-52; 82pp; English.  
 PS  
 XX The invention relates to identifying a candidate p53 pathway modulating  
 CC agent in humans that is referred to in the specification as BGALT (beta-  
 CC 1,3 galactosyltransferase). Also disclosed is a method for diagnosing a  
 CC disease in a patient, by contacting a sample with a probe for B3GALT  
 CC expression, and comparing the results with a control, and determining  
 CC whether the results indicate a likelihood of disease. Methods and  
 CC compositions of the invention are useful for the diagnosis and treatment  
 CC of disorders associated with defects in the p53 pathway, such as cancer  
 CC of the breast, colon, kidneys, lung and ovary. The current sequence  
 CC represents a B3GALT nucleic acid sequence referred to in the disclosure  
 CC of the invention  
 XX  
 SQ Sequence 933 BP; 229 A; 234 C; 240 G; 229 T; 0 U; 1 Other;  
 Query Match 91.9%; Score 928.8; DB 10; Length 933;  
 Best Local Similarity 99.7%; Pred. No. 2e-278;  
 Matches 930; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 79 ATGGCTTTCCGAGATGAGATTGATGATATATCTGCTTCTGGTCTCGGGGCTCTTTGT 138  
 Db 1 ATGGCTTTCCGAGATGAGATTGATGATATATCTGCTTCTGGTCTCGGGGCTCTTTGT 60  
 Qy 139 TTGTATTTTACATGTCAGCTAAATCTTTTCAAAGAACAGTCTTTGTTTACAGAAA 198  
 Db 61 TTGTATTTATAGATGAACAGTCTAAATCTTTCAAAGAACAGTCTTTGTTTACAGAAA 120  
 Qy 199 GACGGAACTTCTTAAAGTCCAGATACAGACTGCAGGACAGACCTCCCTTCTCGTC 258  
 Db 121 GACGGAACTTCTTAAAGTCCAGATACAGACTGCAGGACAGACCTCCCTTCTCGTC 180  
 Qy 259 CTGCTGGTGAACCTCATCCCAAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGG 318  
 Db 181 CTGCTGGTGAACCTCATCCCAAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGG 240  
 Qy 319 GGGAAAGAGAGCGGTGAAGGAAAGCAGCTGAAGACATTTCTTCTCTGGGGACACC 378  
 Db 241 GGGAAAGAGAGCGGTGAAGGAAAGCAGCTGAAGACATTTCTTCTCTGGGGACACC 300  
 Qy 379 AGCAGTGCAGCGGAAACAAAAGAGGTGGACAGGAGCGCAGACAGCGGGACATTATC 438  
 Db 301 AGCAGTGCAGCGGAAACAAAAGAGGTGGACAGGAGCGCAGACAGCGGGACATTATC 360  
 Qy 439 CAGAAGGATTTCTAGACGTCTATTACAATCTGACCCCTGAAGACCATGATGGGCATAGAA 498  
 Db 361 CAGAAGGATTTCTAGACGTCTATTACAATCTGACCCCTGAAGACCATGATGGGCATAGAA 420  
 Qy 499 TGGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAACAGACTTCAGACATGTTTC 558  
 Db 421 TGGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAACAGACTTCAGACATGTTTC 480  
 Qy 559 ATCAATGTTGACTATCTGACTGAACCTGCTTCTGAGAGAAAAACAGAACACCAAGTTTTTC 618  
 Db 481 ATCAATGTTGACTATCTGACTGAACCTGCTTCTGAGAGAAAAACAGAACACCAAGTTTTTC 540

619 ACTGGCTTCTTGAAGAACTCAATGAGTTTCCCATCAGGAGCCATTCCAGCAAGTGGTTTGTG 678  
 541 ACTGGCTTCTTGAAGAACTCAATGAGTTTCCCATCAGGAGCCATTCCAGCAAGTGGTTTGTG 600  
 679 AGTAAATCTGAATATCCGTGGGACAGGTACCCACCAATCTGCTCGGACCGGCTACGTG 738  
 601 AGTAAATCTGAATATCCGTGGGACAGGTACCCACCAATCTGCTCGGACCGGCTACGTG 660  
 739 TTTTCTGGCGAGCTGGCGAGTCAGGTGTACAAATGTCTCCAGAGCGTCCCATACATATAA 798  
 661 TTTTCTGGCGAGCTGGCGAGTCAGGTGTACAAATGTCTCCAGAGCGTCCCATACATATAA 720  
 799 CTGGAAGACGTGTTGTGGGGCTCTGCTCGAAGGCTGAACATCAGATTTGGAGGAGCTC 858  
 721 CTGGAAGACGTGTTGTGGGGCTCTGCTCGAAGGCTGAACATCAGATTTGGAGGAGCTC 780  
 859 CACTCCAGCGACCTTTTTCAGGGGCTTACGCTTCTCGATGCTCTTCAAGAGG 918  
 781 CACTCCAGCGACCTTTTTCAGGGGCTTACGCTTCTCGATGCTCTTCAAGAGG 840  
 919 ATCTGGGCTGCACTTCAATCAAGCTCGGACTCTCTTGGACTCTCTGGACTCTGGAGGCTCTAGAG 978  
 841 ATCTGGGCTGCACTTCAATCAAGCTCGGACTCTCTTGGACTCTCTGGACTCTGGAGGCTCTAGAG 900  
 979 AATTCCTGGGGGGAAGATTGTCCGCTCTGTGA 1011  
 901 AATTCCTGGGGGGAAGATTGTCCGCTCTGTGA 933

RESULT 5  
 ABZ56918  
 ID ABZ56918 standard; DNA; 2629 BP.  
 XX AC ABZ56918;  
 XX DT 04-APR-2003 (first entry)  
 DE B3GALT nucleic acid sequence # SEQ ID 10.  
 XX KW B3GALT; beta-1,3 galactosyltransferase; p53; cancer; breast; colon;  
 XX KW kidney; lung; ovary; gene; ds.  
 XX OS Homo sapiens.  
 XX PN WO200299044-A2.  
 XX PD 12-DEC-2002.  
 XX PF 02-JUN-2002; 2002WO-US017356.  
 XX PR 05-JUN-2001; 2001US-0296076P.  
 XX PR 10-OCT-2001; 2001US-0328605P.  
 XX PR 15-FEB-2002; 2002US-0357253P.  
 XX PA (EXEL-) EXELIXIS INC.  
 XX PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
 XX WPI; 2003-156949/15.  
 XX Identifying p53 pathway modulating agents with B3GALT genes, useful for  
 PT the diagnosis and treatment of disorders associated with defects in the  
 PT p53 pathway, such as cancer of the breast, colon, kidneys, lung and  
 PT ovary.  
 XX Disclosure; Page 53-55; 82pp; English.  
 XX The invention relates to identifying a candidate p53 pathway modulating  
 CC agent in humans that is referred to in the specification as B3GALT (beta-  
 CC 1,3 galactosyltransferase). Also disclosed is a method for diagnosing a  
 CC disease in a patient, by contacting a sample with a probe for B3GALT  
 CC expression, and comparing the results with a control, and determining

CC whether the results indicate a likelihood of disease. Methods and  
 CC compositions of the invention are useful for the diagnosis and treatment  
 CC of disorders associated with defects in the p53 pathway, such as cancer  
 CC of the breast, colon, kidneys, lung and ovary. The current sequence  
 CC represents a B3GALT nucleic acid sequence referred to in the disclosure  
 CC of the invention  
 XX SQ Sequence 2629 BP; 632 A; 665 C; 644 G; 688 T; 0 U; 0 Other;  
 Query Match 91.8%; Score 928.6; DB 10; Length 2629;  
 Best Local Similarity 99.6%; Pred. No. 4.4e-278;  
 Matches 931; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 77 AGATGGCTTTCCCGAAGATGAGATTGATGTATATCTGCTTCTGTTCTGGGGGCTCTTT 136  
 DB 299 AATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCTTCTGTTCTGGGGGCTCTTT 358  
 QY 137 GTTTGTATTAGCATGTACAGTCTAAATCTCTTTCAAGAAACAGTCTTTGTTTACAAGA 196  
 DB 359 GTTTGTATTAGCATGTACAGTCTAAATCTCTTTCAAGAAACAGTCTTTGTTTACAAGA 418  
 QY 197 AAGACGGGAACCTTCTTAAGCTCCAGATACAGACTGAGGAGACACCTCCCTTCTCTCG 256  
 DB 419 AAGACGGGAACCTTCTTAAGCTCCAGATACAGACTGAGGAGACACCTCCCTTCTCTCG 478  
 QY 257 TCCTGCTGGTGACCTCATCCCAAAACAGTTGGCTGAGCGATGGCCATCCGGCAGAGCT 316  
 DB 479 TCCTGCTGGTGACCTCATCCCAAAACAGTTGGCTGAGCGATGGCCATCCGGCAGAGCT 538  
 QY 317 GGGGAAAGAGAGGACCGTGAAGGAAAGAGCTGAAGACATTTCTCTCTCTGGGACCA 376  
 DB 539 GGGGAAAGAGAGGATGTTGAAGGAAAGAGCTGAAGACATTTCTCTCTCTGGGACCA 598  
 QY 377 CCAGCAGTGACGCGGAACAAAGAGGTGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 436  
 DB 599 CCAGCAGTGACGCGGAACAAAGAGGTGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 658  
 QY 437 TCAGAGAGGATTTCTTAGACCTCTATTACAATCTGACCTTGAAGACCATGATGGGCATAG 496  
 DB 659 TCAGAGAGGATTTCTTAGACCTCTATTACAATCTGACCTTGAAGACCATGATGGGCATAG 718  
 QY 497 AATGGTCCATCGCTTTTGTCTCAGCGGGGTTTGTGATGAAAACAGACTCAGACATGT 556  
 DB 719 AATGGTCCATCGCTTTTGTCTCAGCGGGGTTTGTGATGAAAACAGACTCAGACATGT 778  
 QY 557 TCATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAAACAGACACACAGGTTT 616  
 DB 779 TCATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAAACAGACACACAGGTTT 838  
 QY 617 TCATCTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGAGCCATTTCAGCAAGTGGTTTG 676  
 DB 839 TCATCTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGAGCCATTTCAGCAAGTGGTTTG 898  
 QY 677 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGACACCGGCTAG 736  
 DB 899 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGACACCGGCTAG 958  
 QY 737 TGTCTTCTGGCGAGCTGGCGAGTCAGGTGTACATGTTCTCCAAGAGCGTCCCATACATTA 796  
 DB 959 TGTCTTCTGGCGAGCTGGCGAGTCAGGTGTACATGTTCTCCAAGAGCGTCCCATACATTA 1018  
 QY 797 AACTGGAAGACGTGTTGTGGGGCTCTGCTCTCGAAGGCTGAAACATCAGATTGGAGGAGC 856  
 DB 1019 AACTGGAAGACGTGTTGTGGGGCTCTGCTCTCGAAGGCTGAAACATCAGATTGGAGGAGC 1078  
 QY 857 TCCACTCCAGCGGACCTTTTTCAGGGGCTTACGCTTCTCCGATATGCTCTTTCAGGA 916  
 DB 1079 TCCACTCCAGCGGACCTTTTTCAGGGGCTTACGCTTCTCCGATATGCTCTTTCAGGA 1138  
 QY 917 GGATCGGGCTGGCACTTTCATCAAGCTCCGACTCTTTCGACTTCTGCGAGGCTCTAG 976  
 DB 1139 GGATCGGGCTGGCACTTTCATCAAGCTCCGACTCTTTCGACTTCTGCGAGGCTCTAG 1198

QY 977 AGAATTCGGGGGAGAGATTGTCGCTGTCTGA 1011  
 DB 1199 AGAATTCGGGGGAGAGATTGTCGCTGTCTGA 1233

RESULT 6

ABZ56917  
 ID ABZ56917 standard; DNA; 2762 BP.

XX ABZ56917;

DT 04-APR-2003 (first entry)

XX B3GALT nucleic acid sequence # SEQ ID 9.

DE B3GALT; beta-1,3 galactosyltransferase; p53; cancer; breast; colon;  
 KW kidney; lung; ovary; gene; ds.

XX Homo sapiens.

XX WO2000299044-A2.

PN 12-DEC-2002.

XX 02-JUN-2002; 2002WO-US017356.

XX 05-JUN-2001; 2001US-0296076P.

PR 10-OCT-2001; 2001US-0328605P.

PR 15-FEB-2002; 2002US-0357253P.

XX (EXEL-) EXELIXIS INC.

XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;

XX WPI; 2003-156849/15.

XX Identifying p53 pathway modulating agents with B3GALT genes, useful for  
 PT the diagnosis and treatment of disorders associated with defects in the  
 PT p53 pathway, such as cancer of the breast, colon, kidneys, lung and  
 PT ovary.

XX Example 5; Page 52-53; 82pp; English.

XX The invention relates to identifying a candidate p53 pathway modulating  
 CC agent in humans that is referred to in the specification as B3GALT (beta-  
 CC 1,3 galactosyltransferase). Also disclosed is a method for diagnosing a  
 CC disease in a patient, by contacting a sample with a probe for B3GALT  
 CC expression, and comparing the results with a control, and determining  
 CC whether the results indicate a likelihood of disease. Methods and  
 CC compositions of the invention are useful for the diagnosis and treatment  
 CC of disorders associated with defects in the p53 pathway, such as cancer  
 CC of the breast, colon, kidneys, lung and ovary. The current sequence  
 CC represents a B3GALT nucleic acid sequence referred to in an example from  
 CC the invention

XX Sequence 2762 BP; 651 A; 703 C; 679 G; 729 T; 0 U; 0 Other;

XX Query Match 91.8%; Score 928.6; DB 10; Length 2762;

XX Best Local Similarity 99.6%; Pred. No. 4.5e-278;

XX Matches 931; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 77 AGTGGCTTCCCGAGATGAGATGATGATATATCTGCTTCTGGTCTGGGGGCTCTTT 136

DB 432 AAATGGCTTCCCGAGATGAGATGATGATATATCTGCTTCTGGTCTGGGGGCTCTTT 491

QY 137 GTTGTATTTTAGCATGTACAGTCTAAATCTCTTCAAGAACAGTCTCTTTGTGTACAAGA 196

DB 492 GTTGTATTTTAGCATGTACAGTCTAAATCTCTTCAAGAACAGTCTCTTTGTGTACAAGA 551

QY 197 AAGCGGAACTTCTTAAGTCTCCAGATACAGACTGCGAGGAGACACTCTCTCTCTCG 256

DB 552 AAGCGGAACTTCTTAAGTCTCCAGATACAGACTGCGAGGAGACACTCTCTCTCTCG 611

QY 257 TCCTGCTGGTGACCTCATCCACAAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGAGCT 316  
 DB 612 TCCTGCTGGTGACCTCATCCACAAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGAGCT 671  
 QY 317 GGGGAAAGAGAGAGCGGTGAAGGGAAAGCAGCTGAAGACATTTCTTCTCTCTGGGACCA 376  
 DB 672 GGGGAAAGAGAGAGCGGTGAAGGGAAAGCAGCTGAAGACATTTCTTCTCTCTGGGACCA 731  
 QY 377 CCAGCAGTGCAGCGGAAACAAAGAGGTGGACAGGAGAGCCAGGACACACGGGACATTA 436  
 DB 732 CCAGCAGTGCAGCGGAAACAAAGAGGTGGACAGGAGAGCCAGGACACACGGGACATTA 791  
 QY 437 TCCAGAGAGGATTTCTAGACGTCTATTACATCTGACCTCAAGACCATGATGGCATAG 496  
 DB 792 TCCAGAGAGGATTTCTAGACGTCTATTACATCTGACCTCAAGACCATGATGGCATAG 851  
 QY 497 AATGGGTCATCGCTTTTGTCTCAGCGCGGCTTTGTGATGATGATGATGATGATGATGAT 556  
 DB 852 AATGGGTCATCGCTTTTGTCTCAGCGCGGCTTTGTGATGATGATGATGATGATGATGAT 911  
 QY 557 TCATCAATGTTGACTATCTGACTGAATCTGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAG 616  
 DB 912 TCATCAATGTTGACTATCTGACTGAATCTGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAG 971  
 QY 617 TCACTGGCTTCTGAAACTCAATGAGTTTCCATCAGGCAGCCATTCAGCAAGTGGTTTG 676  
 DB 972 TCACTGGCTTCTGAAACTCAATGAGTTTCCATCAGGCAGCCATTCAGCAAGTGGTTTG 1031  
 QY 677 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCACCGGCTACG 736  
 DB 1032 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCACCGGCTACG 1091  
 QY 737 TGTGTTTCTGGGACAGTGGCGAGTCAAGTGTACATGCTTCAAGAGGCTGCCATACATTA 796  
 DB 1092 TGTGTTTCTGGGACAGTGGCGAGTCAAGTGTACATGCTTCAAGAGGCTGCCATACATTA 1151  
 QY 797 AACTGGAAGAGCTGTTGTGGGGCTCTGCTCGGAAAGGCTGAAATCATGATTGGAGGAGC 856  
 DB 1152 AACTGGAAGAGCTGTTGTGGGGCTCTGCTCGGAAAGGCTGAAATCATGATTGGAGGAGC 1211  
 QY 857 TCCACTCCAGCCGACCTTTTCCAGGGGCTTTACGCTTCTCCGTATGCTCTTTCAGGA 916  
 DB 1212 TCCACTCCAGCCGACCTTTTCCAGGGGCTTTACGCTTCTCCGTATGCTCTTTCAGGA 1271  
 QY 917 GGATCGTGGCTGCCACTTTCATCAAGCTCGGACTCTTGGACTACTTGGCAGGCTCTAG 976  
 DB 1272 GGATCGTGGCTGCCACTTTCATCAAGCTCGGACTCTTGGACTACTTGGCAGGCTCTAG 1331  
 QY 977 AGAATTCGGGGGGAAGATTGTCGCTGTCTGA 1011  
 DB 1332 AGAATTCGGGGGGAAGATTGTCGCTGTCTGA 1366

RESULT 7

AAA93875

ID AAA93875 standard; DNA; 2775 BP.

XX AC AAA93875;

XX DT 15-JAN-2001 (first entry)

XX Human beta-1,3 galactose transferase encoding DNA.

XX Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;  
 KW digestive system; ds.

XX Homo sapiens.

XX WO2000050608-A1.

XX 31-AUG-2000.

XX 24-FEB-2000; 2000WO-JP001070.

XX 25-FEB-1999; 99JP-00047571.  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX Narimatsu H, Ishihiki S, Togayachi A, Sasaki K;  
XX WPI; 2000-549409/50.  
XX P-PSDB; AAB93875.  
XX Beta-1,3 galactose transferase and DNA encoding it, useful for synthesis  
XX of type 1 sialyl Lewis, a carbohydrate for treatment of digestive system  
XX cancer.  
XX Claim 5; Page 99-102; 123pp; Japanese.  
XX This invention relates to a polypeptide (I) with beta-1,3 galactose  
XX transferase activity, or variants of (I) comprising amino acid additions,  
XX deletions and/or substitutions. Included in the invention is DNA encoding  
XX all or part of (I); expression vectors containing the DNA, host cells  
XX transformed by the vectors; a method for the preparation of the  
XX polypeptide by culture of the transformants or by expression in the milk  
XX of a transgenic mammal, and antibodies recognising (I). The Beta-1,3  
XX galactose transferase protein transfers galactose by beta-1,3 bonding to  
XX N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as  
XX GlcNAc beta1-3Gal beta1-4Glc) to give Gal beta1-3GlcNAc. The protein and DNA  
XX encoding it are useful for the treatment and diagnosis of cancer of the  
XX digestive system. The present sequence represents Beta-1,3 galactose  
XX transferase encoding DNA  
XX  
XX Sequence 2775 BP; 681 A; 698 C; 669 G; 727 T; 0 U; 0 Other;  
Query Match 91.8%; Score 928.6; DB 3; Length 2775;  
Best Local Similarity 99.6%; Pred. No. 4.5e-278;  
Matches 931; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
77 AGATGGCTTCCCGAAGATGAGTGTATATCTGCTTCTGGTTCGGGGCTCTTT 136  
400 AAATGGCTTCCCGAAGATGAGTGTATATTTGGCTTCTGGTTCGGGGCTCTTT 459  
137 GTTTGTATTTAGCATGTACAGTCTAAATCTTTCAAAGAACAGTCTTTGTACAGA 196  
460 GTTTGTATTTAGCATGTACAGTCTAAATCTTTCAAAGAACAGTCTTTGTACAGA 519  
197 AAGACGGGAATCTCTTAAGCTCCAGATACAGACTGACGAGCAGACCTCCCTTCCTCG 256  
520 AAGACGGGAATCTCTTAAGCTCCAGATACAGACTGACGAGCAGACCTCCCTTCCTCG 579  
257 TCCTGCTGGTGACCTCATCCCAACAAACAGTGTGGCTGAGCGCATGCGCATCCGGCAGAGCT 316  
580 TCCTGCTGGTGACCTCATCCCAACAAACAGTGTGGCTGAGCGCATGCGCATCCGGCAGAGCT 639  
317 GGGGAAAGAGAGAGCGGTGAAGGAAAGCAGCTGAAGACATTTCTCTCTCTGGGACCA 376  
640 GGGGAAAGAGAGAGTGGTGAAGGAAAGCAGCTGAAGACATTTCTCTCTCTGGGACCA 699  
377 CCAGCAGTGCAGCGGAACAAAGAGGTGGACCGAGAGCGCAGCAGCGGAGACATTA 436  
700 CCAGCAGTGCAGCGGAACAAAGAGGTGGACCGAGAGCGCAGCAGCGGAGACATTA 759  
437 TCCAGAGGATTTCTTAGACGCTCTATTAACAATCTGACCTGAAGACCATGATGGGCATAG 496  
760 TCCAGAGGATTTCTTAGACGCTCTATTAACAATCTGACCTGAAGACCATGATGGGCATAG 819  
497 AATGGTCCATCGCTTTTGTCTCAGCGGGCTTTGTGATGAAAGACAGCTCAGACATGT 556  
820 AATGGTCCATCGCTTTTGTCTCAGCGGGCTTTGTGATGAAAGACAGCTCAGACATGT 879  
557 TCATCAATGTTGACTATCTGACTGAATCTTCTGAAGAAAACAGACACACAGGTTTT 616  
880 TCATCAATGTTGACTATCTGACTGAATCTTCTGAAGAAAACAGACACACAGGTTTT 939  
617 TCACTGGGCTTCTGAAACTCAATGAGTTTCCCATCAGGCGAGCCATTCAGCAAGTGGTTG 676

Db 940 TCACTGGGCTTCTGAAACTCAATGAGTTTCCCATCAGGCGAGCCATTCAGCAAGTGGTTG 999  
Qy 677 TCAGTAAATCTGAATATCCGTGGGACAGAGTACCACCAATTCGTCTCCGGCACCGGCTAGG 736  
Db 1000 TCAGTAAATCTGAATATCCGTGGGACAGAGTACCACCAATTCGTCTCCGGCACCGGCTAGG 1059  
Qy 737 TGTTTTCTGGCGACGTGGCGAGTCAAGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 796  
Db 1060 TGTTTTCTGGCGACGTGGCGAGTCAAGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 1119  
Qy 797 AACTGGAAGAGCTTTTGTGGGGCTCTGCTCGAAAGGCTGAACATCAGATTGGAGGAGC 856  
Db 1120 AACTGGAAGAGCTTTTGTGGGGCTCTGCTCGAAAGGCTGAACATCAGATTGGAGGAGC 1179  
Qy 857 TCCACTCCCGACCGACCTTTTTCAGGGGCTTTACGCTTCTCGTATGCTCTTTCAGGA 916  
Db 1180 TCCACTCCCGACCGACCTTTTTCAGGGGCTTTACGCTTCTCGTATGCTCTTTCAGGA 1239  
Qy 917 GGATCGTGGCTGCCACTTTCATCAAGGCTCGGACTCTCTTGGACTTACTGGCAGGCTCTAG 976  
Db 1240 GGATCGTGGCTGCCACTTTCATCAAGGCTCGGACTCTCTTGGACTTACTGGCAGGCTCTAG 1299  
Qy 977 AGAATTCGGGGGGAAGATTGTCGGCTGTCTGA 1011  
Db 1300 AGAATTCGGGGGGAAGATTGTCGGCTGTCTGA 1334  
RESULT 8  
ADK68456  
ID ADK68456 standard; cDNA; 933 BP.  
XX AC ADK68456;  
XX DT 06-MAY-2004 (first entry)  
XX DE Human beta-1,3-galactosyl transferase cDNA SeqID 1.  
XX KW human; gene; ss; saccharide binding protein; maltose binding protein;  
XX KW beta-1,3-galactosyl transferase; inflammation; infectious disease;  
XX KW cancer metastasis suppression; dairy product; antiinflammatory;  
XX KW antimicrobial; cytostatic.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
XX CDS 1..933  
XX FT /\*tag= a  
XX FT /product= "Beta-1,3-galactosyl transferase protein"  
XX PN JP2004016117-A.  
XX PD 22-JAN-2004.  
XX PF 17-JUN-2002; 2002JP-00176132.  
XX PR 17-JUN-2002; 2002JP-00176132.  
XX PA (TOYM ) TOYOBO KK.  
XX DR WPI; 2004-113878/12.  
XX DR P-PSDB; ADK68457.  
XX PT New fusion protein useful in diagnosis of diseases and in manufacture of  
XX PT pharmaceutical products, comprises a saccharide binding protein and beta  
XX PT 1, 3-galactosyl transferase.  
XX PS Example 1; SEQ ID NO 1; 23pp; Japanese.  
XX CC This invention relates to a novel recombinant fusion protein that  
XX CC comprises a saccharide binding protein, in particular a maltose binding  
XX CC protein, and a beta-1,3-galactosyl transferase and an appropriate  
XX CC manufacturing method. Specifically, it refers to fusion protein that can

CC transfer galactose to an N-acetyl glucosamine residue or N-acetyl  
CC glucosamine monosaccharide. The present invention describes a method to  
CC produce this enzyme, cheaply and efficiently, for the diagnosis and  
CC treatment of inflammation, infectious diseases or for cancer metastasis  
CC suppression. Furthermore, it can also be useful for improving dairy  
CC products. Accordingly, compositions exhibit various activities including  
CC antiinflammatory, antimicrobial and cytostatic. This polynucleotide  
CC sequence is the human beta-1,3-galactosyl transferase cDNA of the  
CC invention.

XX SQ Sequence 933 BP; 227 A; 232 C; 241 G; 233 T; 0 U; 0 Other;

Query Match		91.8%;	Score 928.2;	DB 12;	Length 933;
Beat Local Similarity		99.7%;	Pred. No. 3.1e-278;		
Matches 930;		Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;
Qy	79	ATGGCTTTCCCGAAGATGAGATTGATATATCTGCTTCTGGTTCCTGGGGCTCTTTGT	138		
Db	1	ATGGCTTTCCCGAAGATGAGATTGATATATTTGCTTCTGGTTCCTGGGGCTCTTTGT	60		
Qy	139	TGTATTTTAGCATGTACAGTCTAAATCCTTCAAGAACAGTCTTTGTTTACAGAAA	198		
Db	61	TGTATTTTAGCATGTACAGTCTAAATCCTTCAAGAACAGTCTTTGTTTACAGAAA	120		
Qy	199	GACGGAACTTCTTAAAGCTCCAGATACAGACTGACGAGACACCTCCCTTCTCGTC	258		
Db	121	GACGGAACTTCTTAAAGCTCCAGATACAGACTGACGAGACACCTCCCTTCTCGTC	180		
Qy	259	CTGCTGGTGACCTCATCCCAAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGG	318		
Db	181	CTGCTGGTGACCTCATCCCAAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGG	240		
Qy	319	GGGAAAGAGAGAGCGTGAAGGAAAGCAGCTGAAGACATTTCTCTCTGGGACACACC	378		
Db	241	GGGAAAGAGAGAGTGGTGAAGGAAAGCAGCTGAAGACATTTCTCTCTGGGACACACC	300		
Qy	379	AGCAGTCGACGGGAAACAAAAGAGTGCACAGGAGACGACGAGGACGACATTAATC	438		
Db	301	AGCAGTCGACGGGAAACAAAAGAGTGCACAGGAGACGACGAGGACGACATTAATC	360		
Qy	439	CAGAAGGATTTCTTAGAGCTTATTACAACTCTGACCTGAAAGACCATGATGGGCATAGAA	498		
Db	361	CAGAAGGATTTCTTAGAGCTTATTACAACTCTGACCTGAAAGACCATGATGGGCATAGAA	420		
Qy	499	TGGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAACAGACTCAGACATGTTCC	558		
Db	421	TGGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAACAGACTCAGACATGTTCC	480		
Qy	559	ATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAAACAGAACCAACAGGTTTTTC	618		
Db	481	ATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAAACAGAACCAACAGGTTTTTC	540		
Qy	619	ACTGGCTTTCTTGAAACTCAATGAGTTTCCCATCAGGACGCCATTGAGCAAGTGGTTGTC	678		
Db	541	ACTGGCTTTCTTGAAACTCAATGAGTTTCCCATCAGGACGCCATTGAGCAAGTGGTTGTC	600		
Qy	679	AGTAAATCTGAATATCCGTGGGACGAGTACCCACCATTTCTGCTCGGACCGGCTACGTG	738		
Db	601	AGTAAATCTGAATATCCGTGGGACGAGTACCCACCATTTCTGCTCGGACCGGCTACGTG	660		
Qy	739	TTTTCTGGCGAGTGGCGAGTCAGTGTACAAATGTCCTCAAGAGCGTCCCATACATTAAA	798		
Db	661	TTTTCTGGCGAGTGGCGAGTCAGTGTACAAATGTCCTCAAGAGCGTCCCATACATTAAA	720		
Qy	799	CTGGAAGACGTTGTTGTGGGGCTCTGCTCGTCAAGAGGCTGAACATCAGATTTGGAGGAGCTC	858		
Db	721	CTGGAAGACGTTGTTGTGGGGCTCTGCTCGTCAAGAGGCTGAACATCAGATTTGGAGGAGCTC	780		
Qy	859	CACCTCCAGCGACCTTTTTCAGGGGGCTTACGCTTCTCGTATGCTCTTTAGGAGG	918		
Db	781	CACCTCCAGCGACCTTTTTCAGGGGGCTTACGCTTCTCGTATGCTCTTTAGGAGG	840		
Qy	919	ATCGTGGGCTGCGCACTTTCATCAAGCCTCGGACTCTCTTTGGACTACTGCGAGGCTCTAGAG	978		

Db	841	ATCGTGGGCTCGCACTTTCATCAAGCCTCGGACTCTCTTTGGACTACTGCGAGGCTCTAGAG	900
Qy	979	AATTCCCGGGGGGAGATTGTCCGCTGTCTGA	1011
Db	901	AATTCCCGGGGGGAGATTGTCCGCTGTCTGA	933

RESULT 9

ACH79926

ID ACH79926 standard; DNA; 596 BP.

XX AC ACH79926;

XX AC ACH79926;

DT 29-JUL-2004 (first entry)

XX Human genome derived single exon probe #13121.

XX Human; probe; ss; gene expression; single exon probe; microarray;

XX alternative splicing event; genomic alteration.

XX Homo sapiens.

XX US2003194704-A1.

XX 16-OCT-2003.

XX 03-APR-2002; 2002US-00029386.

XX 03-APR-2002; 2002US-00029386.

XX (PENN/) PENN S G.

XX (RANK/) RANK D R.

XX (HANZ/) HANZEL D K.

XX Penn SG, Rank DR, Hanzel DK;

XX WPI; 2004-119264/12.

XX New human genome-derived single exon nucleic acid probes useful for human

gene expression analysis, for identifying or characterizing alternative

splicing events, for assessing genomic alterations or as tools for

surveying tissues.

XX Claim 15; SEQ ID NO 13121; 80pp; English.

CC The invention relates to a nucleic acid probe for measuring human gene

expression, comprising any of the 27,400 fully defined nucleotide

sequences in the specification, or their complements or fragments, and

encoding at least 8 amino acids of any of the 6888 amino acid sequences

fully defined in the specification. The probe is a single exon probe that

hybridises under high stringency conditions to a nucleic acid molecule

expressed in human cells or tissues. Also included are a spatially-

addressable set of single exon nucleic acid probes for measuring human

gene expression (comprising a plurality of single exon nucleic acid

probes cited above, where each of the plurality of probes is separately

and addressably isolatable or amplifiable from the plurality), a single

CC alternative splicing events, in detecting and characterising gross  
CC alterations in the genomic locus that includes their exon, in assessing  
CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
CC or in expressing the ORF-encoded peptide. The present sequence is a human  
CC single exon probe of the invention. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030194704  
XX  
SQ Sequence 596 BP; 149 A; 150 C; 141 G; 156 T; 0 U; 0 Other;  
Query Match 40.1%; Score 405.2; DB 12; Length 596;  
Best Local Similarity 99.3%; Pred. No. 3.1e-115;  
Matches 407; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CCACCTCAGCCTCCTAGCATATAAACTAGACACATCTCTATGTTTGGGTCTAATCATTT 60  
DB 187 CCACCTCAGCCTCCTAGCATATAAACTAGACACATCTCTATGTTTGGGTCTAATCATTT 246  
QY 61 GGATTTTGTCTTTCAGATGCTTTCCGAAAGATGAGATTGATGTATATCTGCTTCTG 120  
DB 247 GGATTTTGTCTTTCAGATGCTTTCCGAAAGATGAGATTGATGTATATTTGCTTCTG 306  
QY 121 GTTCTGGGGCTCTTGTGTTTATTTAGCATGTACAGTCTAAATCCTTTCAAAGACAG 180  
DB 307 GTTCTGGGGCTCTTGTGTTTATTTAGCATGTACAGTCTAAATCCTTTCAAAGACAG 366  
QY 181 TCCCTTTGTTTACAGAAAGCGGAACTTCTTAACTCCAGATACAGACTGCGAGGAG 240  
DB 367 TCCCTTTGTTTACAGAAAGCGGAACTTCTTAACTCCAGATACAGACTGCGAGGAG 426  
QY 241 ACACCTCCCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
DB 427 ACACCTCCCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 486  
QY 301 GCATCCCGCAGACGCTGGGGGAAAGAGAGGAGCGGTGAGGGGAAAGAGCGCTGAAGCATTC 360  
DB 487 GCCATCCCGCAGACGCTGGGGGAAAGAGAGGAGCGGTGAGGGGAAAGAGCGCTGAAGCATTC 546  
QY 361 TTCTCTCTGGGACCAACAGCAGTGCAGCGGAAACAAAAGAGGTGACCA 410  
DB 547 TTCTCTCTGGGACCAACAGCAGTGCAGCGGAAACAAAAGAGGTGACCA 596

RESULT 10  
ABZ56909  
ID ABZ56909 standard; DNA; 1037 BP.  
XX  
AC ABZ56909;  
XX  
DT 04-APR-2003 (first entry)  
XX  
DE B3GALT nucleic acid sequence # SEQ ID 1.  
XX  
KW B3GALT; beta-1,3 galactosyltransferase; p53; cancer; breast; colon;  
KW kidney; lung; ovary; gene; ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 35..1015  
FT /\*tag= a  
FT /product= "B3GALT protein"  
XX  
FN WO200299044-A2.  
XX  
PD 12-DEC-2002.  
XX  
PF 02-JUN-2002; 2002WO-US017356.  
XX  
PR 05-JUN-2001; 2001US-0296076P.  
PR 10-OCT-2001; 2001US-0328605P.  
PR 15-FEB-2002; 2002US-0357253P.

XX (EXEL-) EXELIXIS INC.  
PI Friedman L, Plozman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
XX WPI; 2003-156849/15.  
DR P-PSDB; ABP60061.  
XX  
PT Identifying p53 pathway modulating agents with B3GALT genes, useful for  
PT the diagnosis and treatment of disorders associated with defects in the  
PT p53 pathway, such as cancer of the breast, colon, kidneys, lung and  
PT ovary.  
XX  
PS Example 3; Page 40; 82pp; English.  
XX  
CC The invention relates to identifying a candidate p53 pathway modulating  
CC agent in humans that is referred to in the specification as B3GALT (beta-  
CC 1,3 galactosyltransferase). Also disclosed is a method for diagnosing a  
CC disease in a patient, by contacting a sample with a probe for B3GALT  
CC expression, and comparing the results with a control, and determining  
CC whether the results indicate a likelihood of disease. Methods and  
CC compositions of the invention are useful for the diagnosis and treatment  
CC of disorders associated with defects in the p53 pathway, such as cancer  
CC of the breast, colon, kidneys, lung and ovary. The current sequence  
CC represents a B3GALT nucleic acid sequence referred to in an example from  
CC the invention  
XX  
SQ Sequence 1037 BP; 298 A; 240 C; 213 G; 286 T; 0 U; 0 Other;  
Query Match 16.2%; Score 163.4; DB 10; Length 1037;  
Best Local Similarity 55.7%; Pred. No. 1.1e-39;  
Matches 358; Conservative 0; Mismatches 276; Indels 9; Gaps 2;  
QY 206 ACTTCCTTAAGCTCCAGATACAGACTCGAGCAGACACCTCCCTCTCGTCTGCTGG 265  
DB 225 AATTTCCTTATCAACGAGCCCAATAAATGTGAGAAAAACATTCCTCTTCTGTTATCCTCA 284  
QY 266 TGACCTCATCTCCACAAAACAGTTGGCTGAGCGCATCGGCCATCGGCGAGACGCTGGGGGAAAG 325  
DB 285 TCAGCACCACCTCACAGGAAATTTGATGCCGTGAGCAATCAGAGAGAGCTGGGGGGATG 344  
QY 326 AGAGAGCGGTGAAGGAAAGCAGCTGGAAGACATTCCTCTCTGGGGACCAACAGAGTG 385  
DB 345 AGAACAACTTTAAGGGGATCAAGATAGCCACCTGCTCTCTGGGCAAGATGCTGATC 404  
QY 386 CAGCGGAAACAAAGAGGTGGACAGGAGAGCGGACGACACGGGACACATTATCCAGAGG 445  
DB 405 CTGTTCTCAATCAGATGTTGGAGCAAGAGCGCAAAATCTTCATGATATCATCTGTTGGAG 464  
QY 446 ATTTCTTAGAGCTCTATTACAACTGACCTGAAGACCATGATGGCATAGAAATGGTCC 505  
DB 465 ACTTTATGACTCTCTACCTAACCTTACCCTCAAAACATTAATGGGGATGAGATGGTGG 524  
QY 506 ATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAACAGACTCAGACATGTTTCATCAATG 565  
DB 525 CCACCTTTTGTCAAAGCCCAAGTATGTCATGAAAAACAGACAGCGACATTTTGTAAACA 584  
QY 566 TTGACTATCTGACTGAATGCTTCTGAA-----GAAAAACAGAACACAGGTTTTC 619  
DB 585 TGGACAACTCTTATTATTAATAATTAAGTAAACCTCCACCAAGCCACGAAAGAGTATTTA 644  
QY 620 CTGGCTTCTTGAAACTCAATGATGTTTCCCATCAGCAGCCATTCAGCAAGTGGTTTGCA 679  
DB 645 CTGGCTATGTC-----TTAATGAGACCGATTCGGGATGTCGGCAGTAAAGTATATGTC 701  
QY 680 GTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCCTCTCGGCGACCGCTACGTGT 739  
DB 702 CCAGGATTTGTACCCAGACAGTAACCTACCCACCTTCTGTTGGGGAGCTGCTACATCT 761  
QY 740 TTTCTGGCGAGCTGGCGAGTCAAGTGTATCAATGTTTCCAGAGCGTCCATACATTAAAC 799  
DB 762 TTTGAGCGAGTGTAGCTGAATCTATTTTACAAGACCTCTACTCCACACAAGGCTGCTTACC 821

QY 800 TGAAGACGCTGTTCTTGGGGCTCTGCTCGAAAGGCTGAACAT 842  
|||||  
DB 822 TTGAAGCGTATATGTGGGACTGTCTTCGAAGCTGGGCAT 864  
|||||

## RESULT 11

AAQ67067  
ID AAQ67067 standard; cDNA to mRNA; 1739 BP.

XX AAQ67067;

XX 14-MAR-1995 (first entry)

XX Beta-1,3-galactosyltransferase cDNA.

XX Beta-1,3-galactosyltransferase; enzyme; saccharide chain; pMoPRWM1;  
KW KJM-1 cells; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 676..1656

FT /\*tag= a

FT /product= "Beta-1,3-galactosyltransferase"

XX JP06181759-A.

XX 05-JUL-1994.

XX 16-DEC-1992; 92JP-00336436.

XX 16-DEC-1992; 92JP-00336436.

XX (KYOW ) KYOWA HAKKO KOGYO KK.

XX WPI; 1994-251683/31.

XX P-PSDB; AAR57433.

XX Beta-galactosyl-transferase DNA and protein - useful for prodn. of  
PT saccharide chains.

XX Claim 3; Page 22-24; 47pp; Japanese.

XX This sequence encodes a beta-1,3-galactosyltransferase. This enzyme can  
CC be used to produce physiologically active saccharide chains and variants,  
CC and for improvement of saccharide chains bound to physiologically active  
CC proteins. This cDNA represents a fragment of the plasmid pMoPRWM1 which  
CC was cloned in KJM-1 cells

XX SQ Sequence 1739 BP; 473 A; 407 C; 422 G; 437 T; 0 U; 0 Other;

Query Match 16.2%; Score 163.4; DB 2; Length 1739;  
Best Local Similarity 55.7%; Pred. No. 1.5e-39;  
Matches 358; Conservative 0; Mismatches 276; Indels 9; Gaps 2;

QY 206 ACTTCCTTAAGCTCCAGATACAGACTCCAGGACAGACACCTCCCTCTCGTCTGCTGG 265  
|||||

DB 866 AATTCTTTATCAACGAGGCCCAATAATGTGAGAAAAACATTCCTTTCTGTTATCTCA 925  
|||||

QY 266 TGACCTCATCCCAAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAAG 325  
|||||

DB 926 TCAGACCACTCACAAGGAATTTGATGCCGTGAGCAATCAGAGACGTGGGGGATG 985  
|||||

QY 326 AGAGACCGTGAAGGAAAGCAGCTGAAGACATTTCTCTCTGGGACCAACACAGCAGTG 385  
|||||

DB 986 AGAACAACTTTAAGGGGATCAAGATAGCACCCCTTTCTCTCTGGGCAAGAATGCTGATC 1045  
|||||

QY 386 CAGCGGAACAAAGAGCGTGACAGGAGCCAGCAGCGGACATTTATCCAGAGG 445  
|||||

DB 1046 CTGTTCTCAATCAGATGTGGAGCAAGAGAGCCAAATCTTCATGATATCATCGTGGAGG 1105  
|||||

QY 446 ATTTCTAGACGCTTATTACATCTTGACCTGAAGACCATGATGGGCATAGATGGGTCC 505  
|||||

DB 1106 ACTTTATTGACTCCTACCATAACCTTTACCTCAAAACATTAAATGGGATGAGATGGGTGG 1165  
|||||  
QY 506 ATGCTTTTGTCTCAGGGCGGCTTTTGTGATGAAAAACAGACTCAGACATGTTTCATCAATG 565  
|||||  
DB 1166 CCACTTTTGTTCAAAAGCCAAGTATGTCATGAAAAACAGACGCGACATTTTGTAAACA 1225  
|||||  
QY 566 TTGACTATCTGACTGAACCTGCTTCTGAA-----GAAAAACAGAAACAACAGGTTTTTCA 619  
|||||  
DB 1226 TGGACAATCTTATTATAAATTACTGAAACCTCCACCAAGCCACGAAGAAGGTAATTTA 1285  
|||||  
QY 620 CTGCTCTTTGAAACTCAATGAGTTTCCCATCAGGAGCCATTCAGCAAGTGGTTGTGCA 679  
|||||  
DB 1286 CTGGCTATGTCA---TTAATGGAGGACCGAATTCGGGATGTCOCGACGTAAGTGTATATGC 1342  
|||||  
QY 680 GTAAATCTGAATATCCGTGGGACAGGTACCACATTCTGCTCCGGCACCGGCTACGTGT 739  
|||||  
DB 1343 CCAGGGAATTTGTACCAGACAGTAATCTACCACTTTCTGTTCCGGGACTGGGTACATCT 1402  
|||||  
QY 740 TTTCTGGCGACGTGGCGAGTCAAGTGTACAATGTTCTCAAGAGCGTCCCATACATTAAAC 799  
|||||  
DB 1403 TTTTCAGCCGATGTAGCTGAACATTTTACAAGACCTCACTCCACACAAGGCTGCTTCACC 1462  
|||||  
QY 800 TGAAGACGTTGTTGTGGGCTCTGCTCGAAAGGCTGAACAT 842  
|||||  
DB 1463 TTGAAGACGTATATGTGGGACTGTGCTTCGAAAAGCTGGGCAT 1505  
|||||

## RESULT 12

ABZ56910  
ID ABZ56910 standard; DNA; 2168 BP.

XX AC ABZ56910;

XX 04-APR-2003 (first entry)

XX B3GALT nucleic acid sequence # SEQ ID 2.

XX B3GALT; beta-1,3 galactosyltransferase; p53; cancer; breast; colon;  
KW kidney; lung; ovary; gene; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 352..1332

FT /\*tag= a

FT /product= "B3GALT protein"

XX WO200299044-A2.

XX 12-DEC-2002.

XX 02-JUN-2002; 2002WO-US017356.

XX 05-JUN-2001; 2001US-0296076P.

XX 10-OCT-2001; 2001US-0328605P.

XX 15-FEB-2002; 2002US-0357253P.

XX (EXEL-) EXELIXIS INC.

XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;

XX WPI; 2003-156849/15.

XX P-PSDB; ABP60062.

XX Identifying p53 pathway modulating agents with B3GALT genes, useful for  
PT the diagnosis and treatment of disorders associated with defects in the  
PT p53 pathway, such as cancer of the breast, colon, kidneys, lung and  
PT ovary.

XX Disclosure; Page 41-42; 82pp; English.

XX The invention relates to identifying a candidate p53 pathway modulating  
CC agent in humans that is referred to in the specification as B3GALT (beta-



CC 1,3 galactosyltransferase). Also disclosed is a method for diagnosing a  
CC disease in a patient, by contacting a sample with a probe for B3GALT  
CC expression, and comparing the results with a control, and determining  
CC whether the results indicate a likelihood of disease. Methods and  
CC compositions of the invention are useful for the diagnosis and treatment  
CC of disorders associated with defects in the p53 pathway, such as cancer  
CC of the breast, colon, kidneys, lung and ovary. The current sequence  
CC represents a B3GALT nucleic acid sequence referred to in the disclosure  
CC of the invention

XX SQ Sequence 2168 BP; 641 A; 449 C; 444 G; 634 T; 0 U; 0 Other;  
Query Match 16.0%; Score 161.8; DB 10; Length 2168;  
Best Local Similarity 55.5%; Pred. No. 5.4e-39;  
Matches 357; Conservative 0; Mismatches 277; Indels 9; Gaps 2;  
QY 206 ACTTCCTTAAAGCTCCAGATACAGACTGCAGGACAGACACCTCCCTTCCTGCTGCTGG 265  
Db 542 AATTTCTTATCAGAGCCCAATAATGTGAGAAAAACATTCCTTTCTTGTATTCCTCA 601  
QY 266 TGACCTCATCCCAAAACAGTTGGCTGAGCGCATGGCCATCGGCAGACGCTGGGGGAAAG 325  
Db 602 TCAGCACCACCTCACAAGGAATTTGATGCCGTCAGGCAATCAGAGACGCTGGGGGATG 661  
QY 326 AGAGACCGTGAAGGAAGCAGCTGAAGACATTTCTTCCTGTTGGGACACACGACGATG 385  
Db 662 AGAACAACTTTAAGGGGATCAAGATAGCACCCCTGCTCTCTGGGCAAGATGCTGATC 721  
QY 386 CAGCGGAACAAAGAGCTGACAGGAGAGCCGACAGCGGACACATTCACGAGG 445  
Db 722 CTGTTCTCAATCAGATGTGGAGCAAGAGAGCCAAATCTTCATGATATCATCGTGGAG 781  
QY 446 ATTTCTAGAGCTTATTACATCTGACCCCTGAAGACCATGATGGGCATAGAACTGGTCC 505  
Db 782 ACTTTATGACTCTTACCATACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 841  
QY 506 ATCGCTTTTGTCTCAGCGGGGCTTTGTGATGAAACAGACTCAGACATGTTCAATG 565  
Db 842 CCACCTTTTGTCAAAGCCAAAGTATGTCATGAAACAGACAGCAGCATTTTGTAAACA 901  
QY 566 TTGACTATCTGACTGAATGCTTCTGAA-----GAAAAAGACACACAGGTTTTC 619  
Db 902 TGGCAATCTTATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 961  
QY 620 CTGGCTTCTGAAATCAATGATTTCCATCAGGACGACCATTCAGCAAGTGGTTTGCA 679  
Db 962 CTGGCTATGTCA--TTAATGGAGACCGATTCGGGATGTCGGAGTAATGTTATATGC 1018  
QY 680 GTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCGTCTCGGACACCGGCTACG 739  
Db 1019 CCAGGATTTGTACCAGACAGTAATACCTACCCACCTTTCTGTTGGGGACTGGCTACATCT 1078  
QY 740 TTTCTGGCGAGTGGCGAGTCAGGTGATCAATGTTCCAGAGCGTCCCATACATTAAC 799  
Db 1079 TTTGAGCGATGTAGCTGAATCAATTTTCAAGACCTCACTCCACAGAGGCTGCTTACC 1138  
QY 800 TGGAGAGCTGTTTGTGGGGCTCTCCCTCGAAAGCTGAACAT 842  
Db 1139 TTGAAGAGCTATATGTGGGACTGTGTCTTCGAAGCTGGGCAT 1181

RESULT 13  
AD000390  
ID AD000390 standard; cDNA; 1092 BP.  
XX AC  
XX AD000390;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
XX DE Novel human cDNA sequence #1205.  
XX  
XX ds; gene; anti-inflammatory; dermatological; neuroprotective;  
KW immunomodulator; antibacterial; virucide; antipsoriatic; cytostatic;  
KW

KW gene therapy; vaccine; inflammatory; CNS; immune disorder; cancer;  
KW psoriasis; diabetes; early aging; hormonal imbalance;  
KW ischemic heart disease; ulcerative colitis.  
OS Homo sapiens.  
XX WO2004038003-A2.  
XX PD 06-MAY-2004.  
XX PF 24-OCT-2003; 2003WO-US033947.  
XX  
PR 25-OCT-2002; 2002US-0421061P.  
PR 25-OCT-2002; 2002US-0421080P.  
PR 25-OCT-2002; 2002US-0421552P.  
PR 25-OCT-2002; 2002US-0421614P.  
PR 30-OCT-2002; 2002US-0422177P.  
PR 30-OCT-2002; 2002US-0422178P.  
PR 15-NOV-2002; 2002US-0426355P.  
PR 15-NOV-2002; 2002US-0426384P.  
PR 15-NOV-2002; 2002US-0426394P.  
PR 15-NOV-2002; 2002US-0426430P.  
PR 15-NOV-2002; 2002US-0426916P.  
PR 27-NOV-2002; 2002US-0429224P.  
PR 27-NOV-2002; 2002US-0429275P.  
PR 27-NOV-2002; 2002US-0429302P.  
PR 27-NOV-2002; 2002US-0429326P.  
PR 27-NOV-2002; 2002US-0429651P.  
PR 04-DEC-2002; 2002US-0430645P.  
PR 04-DEC-2002; 2002US-0430651P.  
PR 04-DEC-2002; 2002US-0430657P.  
PR 04-DEC-2002; 2002US-0430663P.  
PR 04-DEC-2002; 2002US-0430668P.  
PR 04-DEC-2002; 2002US-0430684P.  
PR 05-DEC-2002; 2002US-0430937P.  
PR 05-DEC-2002; 2002US-0430965P.  
PR 05-DEC-2002; 2002US-0431458P.  
PR 12-DEC-2002; 2002US-0433251P.  
PR 12-DEC-2002; 2002US-0433500P.  
PR 13-DEC-2002; 2002US-0433316P.  
PR 13-DEC-2002; 2002US-0433318P.  
PR 23-DEC-2002; 2002US-0436236P.  
PR 03-JAN-2003; 2003US-0437914P.  
PR 17-JAN-2003; 2003US-0440820P.  
PR 17-JAN-2003; 2003US-0440821P.  
PR 18-APR-2003; 2003US-0463700P.  
PR 18-APR-2003; 2003US-0463708P.  
PR 18-APR-2003; 2003US-0463716P.  
PR 02-MAY-2003; 2003US-0467199P.  
PR 02-MAY-2003; 2003US-0467201P.  
PR 02-MAY-2003; 2003US-0467203P.  
PR 02-MAY-2003; 2003US-0467230P.  
PR 19-MAY-2003; 2003US-0471306P.  
PR 19-MAY-2003; 2003US-0471336P.  
PR 22-MAY-2003; 2003US-0472420P.  
PR 09-JUN-2003; 2003US-0476609P.  
PR 09-JUN-2003; 2003US-0476621P.  
PR 09-JUN-2003; 2003US-0476632P.  
PR 09-JUN-2003; 2003US-0476641P.  
PR 08-JUL-2003; 2003US-0485217P.  
PR 08-JUL-2003; 2003US-0485218P.  
PR 08-JUL-2003; 2003US-0485224P.  
PR 08-JUL-2003; 2003US-0485325P.  
PR 14-JUL-2003; 2003US-0485359P.  
PR 14-JUL-2003; 2003US-0486446P.  
PR 15-JUL-2003; 2003US-0486480P.  
PR 15-JUL-2003; 2003US-0486891P.  
PR 08-AUG-2003; 2003US-0486960P.  
PR 08-AUG-2003; 2003US-0493341P.  
PR 08-AUG-2003; 2003US-0493370P.







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2005, 12:15:17 ; Search time 210.477 Seconds  
(without alignments)  
7859.644 Million cell updates/sec

Title: US-10-777-828-8

Perfect score: 1011

Sequence: 1 ccactcagctcctcctagcat.....aagatttcgcgtgtctga 1011

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1011	100.0	1011	4	US-09-831-630-8
2	114.8	11.4	1134	3	US-09-482-180A-3
3	78.4	7.8	1420	3	US-09-482-180A-1
4	78.4	7.8	22374	4	US-09-949-016-14938
5	70.2	6.9	1191	3	US-09-459-133-3
6	68.6	6.8	1167	3	US-09-459-133-14
7	67.4	6.7	1464	4	US-09-774-528-430
8	52.8	5.2	1434	2	US-09-055-097-2
9	52.8	5.2	1434	4	US-09-373-902-2
10	52.8	5.2	1640	4	US-09-949-016-5696
11	52.8	5.2	5640	4	US-09-949-016-17438
12	52.8	5.2	8436	4	US-09-949-016-13963
13	46	4.5	125902	4	US-09-949-016-13715
14	44.8	4.4	1532	3	US-09-459-133-1
15	43.4	4.3	1446	3	US-09-459-133-12
16	43.2	4.3	31166	4	US-09-949-016-13464
17	41.4	4.1	7218	1	US-08-233-463-14
18	39.2	3.9	748	4	US-09-270-767-444
19	39.2	3.9	748	4	US-09-270-767-15726
20	39.2	3.9	4079	4	US-09-016-434-1248
21	39.2	3.9	4517	4	US-09-949-016-4573
22	39.2	3.9	4519	4	US-09-023-655-1202
23	39.2	3.9	4519	4	US-09-949-016-650
24	39.2	3.9	4519	4	US-09-492-027-3
25	39.2	3.9	317366	4	US-09-949-016-16001
26	38.6	3.8	17731	4	US-09-949-016-16365
27	38.2	3.8	601	4	US-09-949-016-104020

ALIGNMENTS

RESULT 1

US-09-831-630-8  
; Sequence 8, Application US/09831630  
; Patent No. 6800468  
; GENERAL INFORMATION:  
; APPLICANT: Clausen, Henrik  
; APPLICANT: Amado, Margarita  
; TITLE OF INVENTION: UDP-GALACTOSE: BETA-N-ACETYL-GLUCOSAMINE BETA 1,3  
; FILE REFERENCE: 7188-157  
; CURRENT APPLICATION NUMBER: US/09/831.630  
; CURRENT FILING DATE: 2001-07-12  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 1011  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (79)..(1008)  
US-09-831-630-8

Query Match	100.0%	Score	1011;	DB	4;	Length	1011;
Best Local Similarity	100.0%	Pred. No.	7.5e-313;	Mismatches	0;	Indels	0;
Matches	1011;	Conservative	0;	0;	0;	Gaps	0;
Qy	1	CCACCTCAGCCTCCTAGCATAAACTAGACACATCCTCATCTTTTGGAGGCTTAATCATTT	60				
Db	1	CCACCTCAGCCTCCTAGCATAAACTAGACACATCCTCATCTTTTGGAGGCTTAATCATTT	60				
Qy	61	GGATTTTGTTCCTTTCAGATGGCTTTCCGGAAGATGAGATTGATGATATATCTGCTCTG	120				
Db	61	GGATTTTGTTCCTTTCAGATGGCTTTCCGGAAGATGAGATTGATGATATATCTGCTCTG	120				
Qy	121	GTTCCTGGGGGCTCTTTTGTGTTATTTAGCATGTACAGTCTAAATCTCTTCAAGAACAG	180				
Db	121	GTTCCTGGGGGCTCTTTTGTGTTATTTAGCATGTACAGTCTAAATCTCTTCAAGAACAG	180				
Qy	181	TCCTTTGTTTACAGAAAGACGGCACTTCCTTAAGCTCCAGATACAGCTCGAGGAG	240				
Db	181	TCCTTTGTTTACAGAAAGACGGCACTTCCTTAAGCTCCAGATACAGCTCGAGGAG	240				
Qy	241	ACACCTCCCTTCCTCGTCTGCTGGTGCATCCCAACAAAGTGGCTGAGCGCATG	300				
Db	241	ACACCTCCCTTCCTCGTCTGCTGGTGCATCCCAACAAAGTGGCTGAGCGCATG	300				
Qy	301	GCCATCCGGCAGACGTGGGGGAAAGAGAGACGGTGAAGGAAAGACGCTGAAGACATTC	360				
Db	301	GCCATCCGGCAGACGTGGGGGAAAGAGAGACGGTGAAGGAAAGACGCTGAAGACATTC	360				

```

QY 361 TTCTCTCTGGGACACACAGCAGTGCAGCGAAGAAAGAGAGTGGACAGGAGCCAG 420
Db 361 TTCTCTCTGGGACACACAGCAGTGCAGCGAAGAAAGAGAGTGGACAGGAGCCAG 420
QY 421 CGACAGGGGACATTATCCAGAAGGATTTCTAGACGTCATTATCAATCTCAGCCCTGAAG 480
Db 421 CGACAGGGGACATTATCCAGAAGGATTTCTAGACGTCATTATCAATCTCAGCCCTGAAG 480
QY 481 ACCATGATGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGGCGGTTTGTGATGAAA 540
Db 481 ACCATGATGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGGCGGTTTGTGATGAAA 540
QY 541 ACAGACTCAGACATGTTTCATCAATGTTGACATCTGACTGAACTGCTTCTGAAGAAAAC 600
Db 541 ACAGACTCAGACATGTTTCATCAATGTTGACATCTGACTGAACTGCTTCTGAAGAAAAC 600
QY 601 AGAACACAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGACGCCA 660
Db 601 AGAACACAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGACGCCA 660
QY 720 TTACAGCAAGTGGTTGTTCAGTAAATCTGAAATATCCGTGGGACAGGTACCCACCAATTCG 720
Db 720 TTACAGCAAGTGGTTGTTCAGTAAATCTGAAATATCCGTGGGACAGGTACCCACCAATTCG 720
QY 780 TCCGGCACCGCTACGTGTTTCTGGGACAGTGGGAGTCAAGTGTACAATGTCTCCAG 780
Db 780 TCCGGCACCGCTACGTGTTTCTGGGACAGTGGGAGTCAAGTGTACAATGTCTCCAG 780
QY 840 AGCGTCCCATACATTAAACTCGAAGACGTTTGTGGGCTCTGCTCGAAGGCTGAAC 840
Db 840 AGCGTCCCATACATTAAACTCGAAGACGTTTGTGGGCTCTGCTCGAAGGCTGAAC 840
QY 900 ATCAGATGGAGGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTCTCC 900
Db 900 ATCAGATGGAGGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTCTCC 900
QY 960 GTATGCTCTTCAGGAGATCGTGCCCTGCCACTTCATCAAGCCCTCGGACTCTCTGGAC 960
Db 960 GTATGCTCTTCAGGAGATCGTGCCCTGCCACTTCATCAAGCCCTCGGACTCTCTGGAC 960
QY 1011 TACTGGCAGGCTTAGAGAAATCCCGGGGGAAGATTGTCGCCCTGTCTGA 1011
Db 961 TACTGGCAGGCTTAGAGAAATCCCGGGGGAAGATTGTCGCCCTGTCTGA 1011

```

```

RESULT 2
US-09-482-180A-3
; Sequence 3, Application US/09482180A
; Patent No. 6361985
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Yamamoto, Gayle
; APPLICANT: Gao, Zeren
; APPLICANT: Whitmore, Theodore E.
; APPLICANT: Jaspers, Stephen
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING
; FILE REFERENCE: 98-80
; CURRENT APPLICATION NUMBER: US/09/482,180A
; PRIOR FILING DATE: 2000-01-12
; PRIOR FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: degenerate sequence
; NAME/KEY: variation
; LOCATION: (1)...(1134)

```

```

; OTHER INFORMATION: n is any nucleotide
; NAME/KEY: misc_feature
; LOCATION: (1)...(1134)
; OTHER INFORMATION: n = A,T,C or G
US-09-482-180A-3

```

Query Match 11.4%; Score 114.8; DB 3; Length 1134;  
Best Local Similarity 31.3%; Pred. No. 7.1e-26; Mismatches 358; Indels 3; Gaps 1;

```

Matches 209; Conservative 98; Mismatches 358; Indels 3; Gaps 1;

QY 181 TCCTTTCTTTTACAAGAACGCGGAACTTCCTTAAGCTCCAGATACAGACTGCAGCGAG 240
Db 286 TTYTNACNTAYMGNCAITCYMGNAAYTTYWSNATHYNTYNGARCCNWSNGNTGYWSN 345
QY 241 ACACCTCCCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
Db 346 AARGAYACNTTYTYNTYNTYNGCNATHAARWSNCARCCNGCNCAYGTNGARMGMNGN 405
QY 301 GCCATCCGCGCAGAGCTGGGGA--AAGAGAGGACGCTGAAGGAAAGAGCTGAAGACA 357
Db 406 GCNATHMGNWSNACNTGGGNGMNGNTGGGNGMNGNMGNCARYTNAATYTN 465
QY 358 TTCTTCTCTCTGGGACACACAGCAGTGCAGCGGAAACAAAAGAGGTGGACGAGAGC 417
Db 466 GTNTTYTYNTYNGGNGTNGCNGWSNCCNCNCNCARYTNYTNGCNTAYGARWSN 525
QY 418 CAGCGACACGCGGACATATTCAGAGAGGATTTCTAGACGCTTATTAATCTGACCCCTG 477
Db 526 MNGARTTYGAYGAYATHYTNCAITGCGGAYTTVACNGARGAYTTYTYAAAYTNACNYTN 585
QY 478 AAGACCATGATGGGCATAGAAATGGTCCATCGCTTTTGTCTCAGCGCGGTTTGTGATG 537
Db 586 AARGARYNCAYTNCARMGNTGGGNTGNGCNGTGYCCNCARGNCAYTTTATYTN 645
QY 538 AAAACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAA 597
Db 646 AARGENGAYGAYGAYTNTTYGTCNCAATGTCNCAATGTCNCAATGTCNCAATGTCNCA 705
QY 598 AACAGAAACAGAGTTTTCATCGCTTCTTGAACCTCAATGAGTTTCCATCAGGAGCAG 657
Db 706 GAYCCNGCNCARGAYTYNTYNTYNGGNGAYGTNATHMNCARGCNYTNCNCAAYMGNAA 765
QY 658 CCATTACAGCAAGTGGTTTGTTCAGTAAATCTGAAATCTCCGTGGGACAGGTACCCACCATTC 717
Db 766 ACNARGTNARTAYTYTATCCNCCNWSNATGTAYMNGCNCNCAATYATCCNCCNATY 825
QY 718 TGCTCCGCGCACCGCTACGTTTCTTGGGACGCTGGGAGTCAAGTGTACATGTCTCC 777
Db 826 CGNGGNGGNGGNTAYGTNATGWSNMGNCNACNCTNMNGNMTNCAARGCNATHATG 885
QY 778 AAGAGCGTCCCATACATTAACCTGGAAGACGTTTGTGGGGCTCTGCTGGAAGAGCTG 837
Db 886 GARGAYCNGARYTNTYNTYNTYNTYNTYNTYNTYNTYNTYNTYNTYNTYNTYNTYNTY 945
QY 838 AACATCAG 845
Db 946 GGNVTNWS 953

```

```

RESULT 3
US-09-482-180A-1
; Sequence 1, Application US/09482180A
; Patent No. 6361985
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Yamamoto, Gayle
; APPLICANT: Gao, Zeren
; APPLICANT: Whitmore, Theodore E.
; APPLICANT: Jaspers, Stephen
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING
; FILE REFERENCE: 98-80
; CURRENT APPLICATION NUMBER: US/09/482,180A

```

; CURRENT FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/115,721  
; PRIOR FILING DATE: 1999-01-12  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 1420  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (135)...(1271)  
US-09-482-180A-1

Query Match  
Best Local Similarity 7.8%; Score 78.4; DB 3; Length 1420;  
Matches 299; Conservative 0; Mismatches 346; Indels 3; Gaps 1;  
  
Qy 201 CGGGAATTCCTTAAGCTCCAGATACAGACTGCGAGGAGACACCTCCCTTCCTGCTCCT 260  
Db 440 CCGAAATTTCTATCTTCTGCTGGAGCCTTCAGGCTGTTCCAAAGGATACCTTCTTGCTCCT 499  
  
Qy 261 GCTGTGACCTCATCCCAACAGTTGGCTGAGCGCATGCCATCCGGCAGACGTGGG 320  
Db 500 GGCAATCAAGTCACAGCTGTGTACGTGGAGCGAGTCGGCTATCCGCAACGACGTGGG 559  
  
Qy 321 GAAAGAGAGG---ACGGTGAAGGGAAGCAGCTGAAGACATTTCTTCCTCTGGGGACAC 377  
Db 560 CAGGTGGGGGATGGGCTAGGGCCGCGAGCTGAAGCTGTGTCTCTTCCTAGGGGTGC 619  
  
Qy 378 CAGCAGTCCAGGGGAAACAAAGAGGTGGACGAGGAGCCAGGACACGCGGACATTTAT 437  
Db 620 AGGATCCGCTCCCCAGCCAGCTGCTGGCTATGAGAGTAGGGAGTTTGTATGACATCCT 679  
  
Qy 438 CCAGAAGATTTCTAGAGCTTATTAACTGACCTGAGACCATGATGGGCATAGA 497  
Db 680 CCAGTGGGACTTCACTGAGGACTTCTTCAACCTGACGCTCAAGGAGCTGCACCTGACGCG 739  
  
Qy 498 ATGGGTCCATCGCTTTTGTCTCAGGCGGGCGTTTGTGATGAAACAGACTCAGACATGTT 557  
Db 740 CTGGGTGTGGCTGCTGCTCCCGCCAGCCATTTCTATGCTAAAGGAGATGACGATGCTT 799  
  
Qy 558 CATCAATTTGACTATCTGACTGACTGCTTCTGAAAGAAACAGAAACACGAGTTT 617  
Db 800 TGTCCACGTCGCCAACGTTAGAGTTCTGAGTGGCTGGGACCCAGCCAGGACCTCCT 859  
  
Qy 618 CACTGGCTTCTTGAACCTCAATGATTTCCCATCAGGCGAGCATTCAGCAAGTGGTTGT 677  
Db 860 GGTGGGAGATGTCTATCCGCCAAGCCCTGCCCAACAGGAAACATAAGGTCAAAATACTTCAT 919  
  
Qy 678 CAGTAAATCTGAATATCCGTGGGACAGGTACCCACATTTCTGCTCCGGCAGCCGGCTACGT 737  
Db 920 CCCACCTCAATGTACAGGGCCACCATCTACCCACCTATGCTGTGGGGAGGATATGT 979  
  
Qy 738 GTTTTCTGGCGAGTGGCGAGTCAAGTGTAACAATGTTCTCCAAAGAGCGTCCCATACATPAA 797  
Db 980 CATGTCCAGAGCCACAGTGGCGGCTCCAGCTATCATGGAAGATGCTGAATCTCTC 1039  
  
Qy 798 ACTGGAACAGCTTTTGTGGGCTCTGCTCGAAGGCTGAACATCAG 845  
Db 1040 CATTGATGATGTCTTTTGTGGGTATGTGCTCAGGAGGCTGGGGCTGAG 1087

RESULT 4  
US-09-949-016-14938/C  
; Sequence 14938, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14938  
; LENGTH: 22374  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(22374)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14938

Query Match  
Best Local Similarity 7.8%; Score 78.4; DB 4; Length 22374;  
Matches 299; Conservative 0; Mismatches 346; Indels 3; Gaps 1;  
  
Qy 201 CGGGAATTCCTTAAGCTCCAGATACAGACTGCGAGGAGACACCTCCCTTCCTGCTCCT 260  
Db 21586 CCGAAATTTCTATCTTCTGCTGGAGCCTTCAGGCTGTTCCAAAGGATACCTTCTTGCTCCT 21527  
  
Qy 261 GCTGTGACCTCATCCCAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGGG 320  
Db 21526 GGCCATCAAGTCACAGCTGTGTACGTGGAGCGAGTCGGCTATCCGCAACGACGTGGG 21467  
  
Qy 321 GAAAGAGAGG---ACGGTGAAGGGAAGCAGCTGAAGACATTTCTTCCTCTGGGGACAC 377  
Db 21466 CAGGTGGGGGATGGGCTAGGGCCGCGAGCTGAAGCTGTGTTCTCTTAGGGGTGGC 21407  
  
Qy 378 CAGCAGTGCAGCGGGAACAAAGAGGTGGACCGAGGAGCCAGGACACGCGGACATTTAT 437  
Db 21406 AGGATCCGCTCCCGCCAGCCAGCTGCTGGCTATGAGAGTAGGGAGTTTGTATGACATCCT 21347  
  
Qy 438 CCAGAAGGATTTCTAGAGCTTATTAACTGACCTGAGACCATGATGGGCATAGA 497  
Db 21346 CCAGTGGGACTTCACTGAGGACTTCTTCAACCTGACGCTCAAGGAGCTGCACCTGCAGCG 21287  
  
Qy 498 ATGGGTCCATCGCTTTTGTCTCAGGCGGGGTTTGTGATGAAAAACAGACTCAGACATGTT 557  
Db 21286 CTGGGTGTGGCTGCTGCTCCCGCCAGGCCATTTCTATGCTAAAGGAGATGACGATGCTT 21227  
  
Qy 558 CATCAATTTGACTATCTGACTGACTGCTTCTGAAAGAAACAGAAACACGAGTTT 617  
Db 21226 TGTCCACGTCGCCAACGTTAGAGTTCTCTGGATGGCTGGGACCCAGCCAGGACCTCCT 21167  
  
Qy 618 CACTGGCTTCTTGAACCTCAATGATTTCCCATCAGGCGAGCATTCAGCAAGTGGTTGT 677  
Db 21166 GGTGGGAGATGTCTATCCCGAAGCCCTGCCCAAGGAAACATAAGGTCAAAATACTTCAT 21107  
  
Qy 678 CAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCAGCCGGCTACGT 737  
Db 21106 CCCACCTCAATGTACAGGGCCACCCACTACCCACCTATGCTGTGGGGAGGATATGT 21047  
  
Qy 738 GTTTTCTGGCGAGCTGGCGAGTCAAGTGTAACAATGTTCTCCAAAGAGCGTCCCATACATPAA 797  
Db 21046 CATGTCCAGAGCCACAGTGGCGGCTCCAGGCTATCATGGAAGATGCTGAATCTTCTCC 20987  
  
Qy 798 ACTGGAAGAGCTGTTTGTGGGCTCTGCTCGAAGGCTGAACATCAG 845  
Db 20986 CATTGATGATGTCTTTTGTGGGTATGTGCTCAGGAGGCTGGGGCTGAG 20939

RESULT 5  
US-09-459-133-3  
; Sequence 3, Application US/09459133  
; Patent No. 6416988  
; GENERAL INFORMATION:

```

; APPLICANT: Conklin, Darrell C.
; APPLICANT: Yamamoto, Gayle
; APPLICANT: Jaspers, Stephen R.
; APPLICANT: Gao, Zeren
; TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS
; FILE REFERENCE: 98-77
; CURRENT APPLICATION NUMBER: US/09/459,133
; CURRENT FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: 60/111,697
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1191
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1191)
; OTHER INFORMATION: n = A,T,C or G
US-09-459-133-3

```

Query Match	6.9%	Score 70.2	DB 3	Length 1191
Best Local Similarity	28.4%	Pred. No. 1.4e-11		
Matches 212	Conservative 95	Mismatches 427	Indels 12	Gaps 3
QY	247	CCCTTCCTCGTCTGCTGCTGACTCATCCACAAGTTCGCTGAGCGCATCGGCATC	306	
DB	442	CCNTAYTNTYNTYNTGCGNTNARWSNGARCCNGNGMNTTYCGNGARMGNCARGCGNTN	501	
QY	307	CGGCAGACGTGGGGGAAAGAGAGGACGGTGAAGGGGAAAGACGCTGAAGACATTCCTCCCTC	366	
DB	502	MNGARACNTGGGGNSNCCNGCNCNGGNA THMGNTYNTNTTYTNTYNTGNGWSNCCN	561	
QY	367	CTGGGACACACAGCAGCTGCAGCGGAAAACAAAGAGGTGGACACGAGAGCCAGCGACAC	426	
DB	562	GTNGNGARGCNGCNGCNGAYYNTGAYWS --- NYTNGTNGCNTGGGARWSNMGMNGTAY	618	
QY	427	GGGGACATTATCAGAAGGATTTCCTAGAGCTCTATTACAATCTGACCCCTGAAGACCATG	486	
DB	619	WSNGAYTNTYNTYNTGGGAYTYYTNGAYGTNCCNTTYAAYCARACNYTNAARGAYTNT	678	
QY	487	ATGGGCATAGAAATGGGTCCATCGCTTTTGTCCTCAGCGCGCGTTTGTGTGATGAAAAACAGAC	546	
DB	679	YTNTYNTGNCNTGGYTNNGNMGNCAVTGYCCNACNGTNSNTTYGNTYTMNGCNCAR	738	
QY	547	TCAGACATGTTTCATCAATGTTTGATCTACTCTCACTGAACCTGCT-----TC TGAAGAAAAAC	600	
DB	739	GAYGAYCNYTGTNCAYACNCCNGCNYNTYTNGCNCAYTNNMGNCNYTNCNCNGCNC	798	
QY	601	AGAACCAACAGGTTTTTCACTGGCTTCTTGAACCTCAATGAGTTCCTCCATCAGGCAGCCCA	660	
DB	799	WSNGCNGWSNTYNTAYTNTGNGARGTNTTYACNARGCNATGCCNYTNMGNAECCN	858	
QY	661	TTCAGCAAGTGGTTTGTTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATCTGTC	720	
DB	859	GGNGGCCNTTYTAYGTNCCNGARWSNTTYTTYGARGNG --- GNTAYCCNGCNTAYGCN	915	
QY	721	TCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACATGCTCCAAG	780	
DB	916	WSNGNGNGNGNTAYGTNATHGNCNGNMGNNTYTNGCNCNTGGYTYTNNMGNCNGCNGCNC	975	
QY	781	AGCGTCCCATACATTAACCTGGAAGACGTGTTGTGGGGCTCTGCGCTCGAAAGCGCTGAAC	840	
DB	976	MNGTNGCNCNTTYCCNTTYGARGAYGTNTAYACNGENYNTGYATHMGNGCNYTNGN	1035	
QY	841	ATCAGATTGAGAGCTCCACTCCAGCCGACCTTTTTTTCAGGGGGCTTACGCTCTCTCC	900	
DB	1036	YTNGTNCCNARGCNAYCCNGGNTTYTNTACNCGNTTGGCNGCNGAYMGNACNGCNGAY	1095	
QY	901	GTATGCCTCTTCAGGAGGATCGTGGCGCTGCCACTTTCATCAAGCCTCGGACTCTCTTGGAC	960	

```

Db      1096 CATTGYCGNTTYMGNAAYTNYTNYTGTTGNGCCNCTYTGNGCNCARGCNWSNATHMGN 1155
          ||: ||::|| | | | | | | | | | | | | | | : || | | : |
Qy      961 TACTGGCAGGCTCTAGAGAATCCCCG 986
          ||| | | : | | | | | | | | | | | | | | : |
Db      1156 YTTGTGAARCARYTNCARGAYCCNMG 1181
          ||| | | : | | | | | | | | | | | | | | : |

RESULT 6
US-09-459-133-14
; Sequence 14, Application US/09459133
; Patent NO. 6416988
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Yamamoto, Gayle
; APPLICANT: Jaspers, Stephen R.
; APPLICANT: Gao, Zeren
; TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS
; FILE REFERENCE: 98-77
; CURRENT APPLICATION NUMBER: US/09/459,133
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: 60/111,697
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 1167
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: degenerate sequence
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1167)
; OTHER INFORMATION: n = A,T,C or G
US-09-459-133-14

```

[illegible]









Db 9845 CGTCTCAACAACAACAACACACCTTATATGGCATTCAGGTGGGTAACCTAGTTTG 9786  
Qy 516 TCCTCAGCGGGTGTGTGATGAACAAGACTAGACATGTTCATCAATGTGTGACTATCT 575  
Db 9785 CCCCATTGCCAACTACATCATGACAAACAGACACTTATGTGTTTCATCAATACTGGCAATTT 9726  
Qy 576 GACTGAACCTGCTCTGAAGAAACAGAACACAGGTTTTTTCATGCTTCTTGAACACT 635  
Db 9725 AGTGAGTATCTTTTAAACTTAAACCACTCAGAGAAGTTTTTTCACAGGTTATCCCTTAAT 9666  
Qy 636 CAATGAGTTTCCCATCAG 653  
Db 9665 TCATAATATTCCCATAG 9648

## RESULT 14

US-09-459-133-1  
; Sequence 1, Application US/09459133  
; Patent No. 6416988  
; GENERAL INFORMATION:  
; APPLICANT: Konklin, Darrell C.  
; APPLICANT: Yamamoto, Gayle  
; APPLICANT: Jaspers, Stephen R.  
; APPLICANT: Gao, Zeren  
; TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS  
; FILE REFERENCE: 98-77  
; CURRENT APPLICATION NUMBER: US/09/459,133  
; CURRENT FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: 60/111,697  
; PRIOR FILING DATE: 1998-12-10  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 1532  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (101)...(1294)  
US-09-459-133-1

Query Match 4.4%; Score 44.8; DB 3; Length 1532;  
Best Local Similarity 49.3%; Pred. No. 0.0022;  
Matches 146; Conservative 0; Mismatches 147; Indels 3; Gaps 1;  
Qy 247 CCCTTCCTCGTCTGCTGGTGACCTCATCCCAACACAGTTGGCTGAGCGCATGGCCATC 306  
Db 542 CCTTACTCTGTTGGCCGTCAAGTCAGAACAGGGCGCTTTGCAGAACGACAGGCCGTG 601  
Qy 307 CGGCAGACGTGGGGAAAGAGAGGACGGTGAAGGGAAAGAGCTGAAGACATTTCTTCCTC 366  
Db 602 AGAGAGACGTGGGGCAGTCCAGCTCCAGGGATCCGGCTGCTTCTCTAGGGTCTCCG 661  
Qy 367 CTGGGGACACACAGCAGTGCAGCGGAAACAAAGAGTGGACAGGAGCAGCGACAC 426  
Db 662 GTGGGTGAGCGGGCCCTGACCTAGACTCACTA---GTGGCATGGGAGAGCGCTCGCTAC 718  
Qy 427 GGGGACATTATCCAGAAGGATTTCTAGACGTCTATTACAATCTGACCCCTGAAGACCATG 486  
Db 719 AGTGACCTGCTGCTCTGGGACTTCTCGACGTCCCATTTCAACAGACGCTCAAGACCTG 778  
Qy 487 ATGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGCGGGCTTTGTGATGAAC 542  
Db 779 CTGCTGCTGGCCTGGCTGGCGCGCACTGCCCCCGTGGTGTGTTTGTCTTGGAGC 834

## RESULT 15

US-09-459-133-12  
; Sequence 12, Application US/09459133  
; Patent No. 6416988  
; GENERAL INFORMATION:  
; APPLICANT: Konklin, Darrell C.

; APPLICANT: Yamamoto, Gayle  
; APPLICANT: Jaspers, Stephen R.  
; APPLICANT: Gao, Zeren  
; TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS  
; FILE REFERENCE: 98-77  
; CURRENT APPLICATION NUMBER: US/09/459,133  
; CURRENT FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: 60/111,697  
; PRIOR FILING DATE: 1998-12-10  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 12  
; LENGTH: 1446  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (107)...(1273)  
US-09-459-133-12

Query Match 4.3%; Score 43.4; DB 3; Length 1446;  
Best Local Similarity 49.1%; Pred. No. 0.006;  
Matches 144; Conservative 0; Mismatches 146; Indels 3; Gaps 1;  
Qy 247 CCCTTCCTCGTCTGCTGGTGACCTCATCCCAACACAGTTGGCTGAGCGCATGGCCATC 306  
Db 524 CCTTACTTGTCTGCTGGCTGTCAATCAGAACAGGACACTTTGCAGCACGCGAGCTGTG 583  
Qy 307 CGGCAGACGTGGGGAAAGAGAGGACGGTGAAGGGAAAGAGCTGAAGACATTTCTTCCTC 366  
Db 584 AGGAGACCTGGGGCAGCCCACTTGTCTGGACCCCGTGTCTTCTCTGCTGGGTCCCCC 643  
Qy 367 CTGGGGACCAACAGCAGTGCAGCGGAAACAAAGAGTGGACAGGAGAGCAGCGACAC 426  
Db 644 CTAGG---AATGGGGGGCCTGACTTAAGATCACTGCTGACGTGGGAAAGCCGCGCTAT 700  
Qy 427 GGGGACATTATCCAGAAGGATTTCTTAGACGTCTATTACAATCTGACCCCTGAAGACCATG 486  
Db 701 GGTGACCTACTGCTCTGGGACTTCTCTGGATGTTCCCTTACAAACCGGACACTCAAGGACTG 760  
Qy 487 ATGGGCATAGAAATGGGTCCCATCGCTTTTGTCTCAGGCGGGCTTTGTGATGAA 539  
Db 761 CTGCTGCTGACCTGGCTGGCAGCCACCACTGCCCGATGTCAATTTTGTCTCTGCA 813

Search completed: September 22, 2005, 17:01:56  
Job time : 213.477 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2005, 07:38:28 ; Search time 525.906 Seconds  
(without alignments)  
10595.706 Million cell updates/sec

Title: US-10-777-828-8\_COPY\_1\_115

Perfect score: 115

Sequence: 1 ccactcagctcctcatgcat.....gagattgatgtatctgcc 115

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_on.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	115	100.0	1011	6	BD228354 UDP-galac
2	113.4	98.6	1360	9	AB041412 Gorilla g
3	113.4	98.6	1565	9	AB041413 Homo sapi
4	113.4	98.6	1576	9	AB041416 Homo sapi
5	113.4	98.6	10562	6	E38420 Novel polyp
6	113.4	98.6	170121	9	AF064860 Homo sapi
7	113.4	98.6	340000	9	HS21C080 Homo sapi
8	111.8	97.2	1566	9	AB041414 Pan trogl
9	111.8	97.2	1579	9	AB041417 Pongo pyg
10	111.8	97.2	192219	9	RP430021119 Pan trogl
11	110.2	95.8	1570	9	AB041415 Pan panis
12	58.2	50.6	170108	2	AC150794 Bos tauru
13	37.2	32.3	243047	2	AC126579 Rattus no
14	37	32.2	933	9	HSA6078 Homo sapi
15	36.8	32.0	131395	9	AC084740 Homo sapi
16	35.8	31.1	2494	6	CQ731786 Sequence
17	35.8	31.1	2762	9	AB020337 Homo sapi
18	35.8	31.1	2775	6	E38419 Novel polyp
19	35.8	31.1	2921	9	AY372061 Homo sapi

20	35.4	30.8	933	9	AF145784 Homo sapi
21	34.4	29.9	199196	2	AC130445 Rattus no
22	34.4	29.9	285998	2	AC127143 Rattus no
23	34.4	29.9	292664	2	AC103298 Rattus no
24	34.2	29.7	158959	2	AC122094 Rattus no
25	34.2	29.7	226295	2	AC119337 Rattus no
26	34	29.6	176814	2	AC121405 Rattus no
27	33.4	29.0	65608	9	AL450243 Human DNA
28	33.4	29.0	123229	2	AC102919 Mus muscu
29	33.4	29.0	181618	2	AL442129 Homo sapi
30	33.4	29.0	209860	2	AC151269 Mus muscu
31	33	28.7	78025	9	HS780M13 Human DNA
32	32.8	28.5	194018	2	AC138568 Gallus ga
33	32.2	28.0	1500	8	AB096162 Zostera c
34	32.2	28.0	161699	9	AC107300 Homo sapi
35	32.2	28.0	167357	2	AC026264 Homo sapi
36	32.2	28.0	168068	9	AL445668 Human DNA
37	32.2	28.0	191038	10	AC124719 Mus muscu
38	32	27.8	110001	9	AC112509 Homo sapi
39	32	27.8	167744	2	AC083941 Homo sapi
40	32	27.8	195269	6	AX417035 Sequence
41	32	27.8	195334	2	AC113657 Rattus no
42	32	27.8	200028	9	AC092680 Homo sapi
43	32	27.8	202354	9	AC006371 Homo sapi
44	32	27.8	333050	1	AL596168 Listeria
45	32	27.8	349980	6	AX417044 Sequence

ALIGNMENTS

RESULT 1	BD228354	1011 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	BD228354	UDP-galactose: beta-N-acetyl-glucosamine beta 1,3 galactosyltransferases, beta 3 Gal-T5.			
DEFINITION	BD228354.1	GI:33038124			
ACCESSION	BD228354				
VERSION	JP 2002530071-A/8.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 1011)				
AUTHORS	Clausen,H. and Amado,M.				
TITLE	UDP-galactose: beta-N-acetyl-glucosamine beta 1,3 galactosyltransferases, beta 3 Gal-T5				
JOURNAL	Patent: JP 2002530071-A 8 17-SEP-2002;				
COMMENT	HENRIK CLAUSEN				
	OS Homo sapiens (human)				
	PN JP 2002530071-A/8				
	PD 17-SEP-2002				
	PF 11-NOV-1999				
	PR 13-NOV-1998				
	PC HENRIK CLAUSEN, MARGARIDA AMADO				
	CC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/10, C12N15/00, C12N5/00				
	CC UDP-galactose: beta-N-acetyl-glucosamine beta 1,3 CC				
	CC galactosyltransferases,				
	CC beta 3 Gal-T5				
	FH Key				
	FT CDS				
	Location/Qualifiers				
	1..1011				
	/organism="Homo sapiens"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:9606"				

FEATURES

Query Match 100.0%; Score 115; DB 6; Length 1011;  
Best Local Similarity 100.0%; Pred. No. 5.6e-25;  
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN

```
Qy 1 CCACCTCAGCTCCTAGCATAAACACACATCCTCATGCTTTTGAGGCTCAATCATT 60
Db 1 CCACCTCAGCTCCTAGCATAAACACACATCCTCATGCTTTTGAGGCTCAATCATT 60

Qy 61 GGATTTTGTCTTTCAGATGCGCTTTCCGGAAGATGAGATTGATGATATCTGCC 115
Db 61 GGATTTTGTCTTTCAGATGCGCTTTCCGGAAGATGAGATTGATGATATCTGCC 115

RESULT 2
AB041412 1360 bp DNA linear PRI 17-OCT-2000
LOCUS Gorilla gorilla betal,3-Galt 5 gene for UDP-Gal:GlcNAC
DEFINITION betal,3-galactosyltransferase 5, partial cds.
ACCESSION AB041412
VERSION AB041412.1 GI:7593018
KEYWORDS
SOURCE
ORGANISM Gorilla gorilla (gorilla)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE 1 (bases 1 to 1360)
AUTHORS Liu, Y. and Saitou, N.
JOURNAL Published Only in DataBase (2000)
TITLE Direct Submission
AUTHORS Liu, Y. and Saitou, N.
JOURNAL Submitted (11-APR-2000) Naruya Saitou, National Institute of
Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima,
Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp,
URL:http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790,
Fax:81-559-81-6789)
FEATURES
Location/Qualifiers
1..1360
/organism="Gorilla gorilla"
/mol_type="genomic DNA"
/isolate="#085"
/db_xref="taxon:9593"
/note="human sequence used for primer design based on Acc#
AB020337"
<1..466
/number=3
467..1360
/gene="betal,3-Galt 5"
CDS 467..>1360
/gene="betal,3-Galt 5"
/db_xref="GI:7593019"
/codon_start=1
/product="UDP-Gal:GlcNAC betal,3-galactosyltransferase 5"
/protein_id="BAA9497.1"
/db_xref="GI:7593019"
/translation="MAFFKRLMYICLLVLGALCLFYSMYSLNPFKEQSFYVKDGNF
LKLPDTCRTPPPFLVLLVTSSSHQLAERMAIROTGWKERTVKGKQLKTFLLGTTSS
AAETKEVDQSQRHGDIIQKDFLDVYNYLTKTMWGLEWHRFCPOAFAVMKTDSDMF
INVDYLLKLNKRTTRFFTGKLEFFIRQPFKSFYKSKSYPDWDRYPFPCSGTG
YVFGSDVASQYVNVSEVPYIKLEDVFGVGLCLERLNRLBELHSQTPFFPGGLRFSVC
LFRIRVACHFIKPTLLIDYQA"
467..>1360
/gene="betal,3-Galt 5"
/number=4
exon

ORIGIN
Query Match 98.6%; Score 113.4; DB 9; Length 1360;
Best Local Similarity 99.1%; Pred. No. 1.8e-24;
Matches 114; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCACCTCAGCTCCTAGCATAAACACACATCCTCATGCTTTTGAGGCTCAATCATT 60
Db 389 CCACCTCAGCTCCTAGCATAAACACACATCCTCATGCTTTTGAGGCTCAATCATT 448

Qy 61 GGATTTTGTCTTTCAGATGCGCTTTCCGGAAGATGAGATTGATGATATCTGCC 115
Db 449 GGATTTTGTCTTTCAGATGCGCTTTCCGGAAGATGAGATTGATGATATCTGCC 503
```

```
RESULT 3
AB041413 1565 bp DNA linear PRI 13-APR-2000
LOCUS Homo sapiens betal,3-Galt 5 gene for UDP-Gal:GlcNAC
DEFINITION betal,3-galactosyltransferase 5, partial cds.
ACCESSION AB041413
VERSION AB041413.1 GI:7593020
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1565)
AUTHORS Liu, Y. and Saitou, N.
JOURNAL Silver Project
TITLE Published Only in DataBase (2000)
REFERENCE 2 (bases 1 to 1565)
AUTHORS Liu, Y. and Saitou, N.
JOURNAL Direct Submission
TITLE Submitted (11-APR-2000) Naruya Saitou, National Institute of
Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima,
Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp,
URL:http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790,
Fax:81-559-81-6789)
FEATURES
Location/Qualifiers
1..1565
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="human-NR"
/db_xref="taxon:9606"
/note="human sequence used for primer design based on Acc#
AB020337"
<1..27
/number=3
28..674
/number=3
675..1565
/gene="betal,3-Galt 5"
675..>1565
/gene="betal,3-Galt 5"
/codon_start=1
/product="UDP-Gal:GlcNAC betal,3-galactosyltransferase 5"
/protein_id="BAA9498.1"
/db_xref="GI:7593021"
/translation="MAFFKRLMYICLLVLGALCLFYSMYSLNPFKEQSFYVKDGNF
LKLPDTCRTPPPFLVLLVTSSSHQLAERMAIROTGWKERTVKGKQLKTFLLGTTSS
AAETKEVDQSQRHGDIIQKDFLDVYNYLTKTMWGLEWHRFCPOAFAVMKTDSDMF
INVDYLLKLNKRTTRFFTGKLEFFIRQPFKSFYKSKSYPDWDRYPFPCSGTG
YVFGSDVASQYVNVSKSPYIKLEDVFGVGLCLERLNRLBELHSQTPFFPGGLRFSVC
LFRIRVACHFIKPTLLTGTGR"
675..>1565
/gene="betal,3-Galt 5"
/number=4
exon

ORIGIN
Query Match 98.6%; Score 113.4; DB 9; Length 1565;
Best Local Similarity 99.1%; Pred. No. 1.8e-24;
Matches 114; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCACCTCAGCTCCTAGCATAAACACACATCCTCATGCTTTTGAGGCTCAATCATT 60
Db 597 CCACCTCAGCTCCTAGCATAAACACACATCCTCATGCTTTTGAGGCTCAATCATT 656

Qy 61 GGATTTTGTCTTTCAGATGCGCTTTCCGGAAGATGAGATTGATGATATCTGCC 115
Db 657 GGATTTTGTCTTTCAGATGCGCTTTCCGGAAGATGAGATTGATGATATCTGCC 711

RESULT 4
AB041416 1576 bp DNA linear PRI 17-OCT-2000
LOCUS
```

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 10562)  
AUTHORS Narimatsu,H., Isehiki,S., Togayauchi,A. and Sasaki,K.  
TITLE Novel polypeptide  
JOURNAL KYOWA HAKKO KOGYO CO LTD  
COMMENT OS Homo sapiens (human)  
PN JP 2000245464-A/2  
PD 12-SEP-2000  
PF 25-FEB-1999 JP 1999047571  
PR  
PI HISASHI NARIMATSU, SOICHIRO ISSHIKI, AKIRA TOGAYAUCHI, PI  
KATSUTOSHI SASAKI  
PC C12N15/09,A01K67/027,C12N1/21,C12N5/10,C12N9/10,C12P19/00, PC  
C12P21/02,  
PC C12P21/08,C12Q1/68,G01N33/53//C12N1/21,C12R1/185), (C12N5/10,  
PC C12R1/91),  
PC (C12P21/02,C12R1/185), (C12P21/02,C12R1/91), C12N5/00,C12N5/00,  
PC (C12N5/00,C12R1/91)  
CC  
FH Key Location/Qualifiers  
FT promoter (1)..(5000)  
FT exon (5001)..(5140)  
FT exon (5001)..(5273)  
FT exon (5459)..(5567)  
FT exon (7427)..(7586)  
FT exon (8234)..(10562).  
FEATURES  
source  
1..10562  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 98.6%; Score 113.4; DB 6; Length 10562;  
Best Local Similarity 99.1%; Pred. No. 1.9e-24;  
Matches 114; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CCACCTCAGCCTCCTAGCATAAACTAGACACATCCTCATGCTTTTGAGTCTAATCAT 60  
DB 8156 CCACCTCAGCCTCCTAGCATAAACTAGACACATCCTCATGCTTTTGAGTCTAATCAT 8215  
QY 61 GGATTTTGTCTTTTCAGATGCTTTCCGGAAGATGAGATTCATGTATATCTGCC 115  
DB 8216 GGATTTTGTCTTTTCAGATGCTTTCCGGAAGATGAGATTCATGTATATTTGCC 8270  
RESULT 6  
AF064860  
LOCUS  
DEFINITION Homo sapiens chromosome 21 clone PAC 70124 map 21q22.3, complete  
sequence.  
ACCESSION AF064860  
VERSION AF064860.2 GI:18958624  
KEYWORDS HTG; HTGS DRAFT.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 170121)  
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,  
Park,H.-S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.-K., Seoda,E.,  
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,  
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,  
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M.B., Schudy,A.,  
Zimmermann,W., Rosenthal,A., Kudoh,J., Kawasaki,K., Asakawa,S.,  
Shintani,A., Sasaki,T., Nagamine,K., Mitsuayama,S.,  
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordtsick,G.,  
Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,  
Reichelt,J., Kauer,G., Blocker,H., Ramser,J., Beck,A., Klages,S.,  
Hennig,S., Risseilmann,L., Dandand,E., Haaf,T., Wehrmeyer,S.,

Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H.,  
Reinhardt,R. and Yaspo,M.Laure  
The DNA sequence of human chromosome 21  
Nature 405 (6784), 311-319 (2000)

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

2 (bases 1 to 170121)  
Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,  
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.  
Direct Submission  
Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany  
3 (bases 1 to 170121)  
Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,  
Weber,J., Schattevoy,R., Yaspo,M.-L., Rosenthal,A., Yaspo,M.-L. and  
Rosenthal,A.  
Direct Submission  
Submitted (27-FEB-2002) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany  
4 (bases 1 to 170121)  
Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,  
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.  
Direct Submission  
Submitted (05-MAR-2002) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany  
On Feb 27, 2002 this sequence version replaced gi:3171153.

## COMMENT

## FEATURES

## source

```
1..170121
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="21"
    /map="21q22.3"
    /clone="PAC 70124"
```

## ORIGIN

Query Match 98.6%; Score 113.4; DB 9; Length 170121;  
Best Local Similarity 99.1%; Pred.No.2.1e-24;  
Matches 114; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCACCTCAGCCTCCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGCTCAATCAT 60  
Db 89970 CCACCTCAGCCTCCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGCTCAATCAT 90029  
Qy 61 GGATTTTGTCTTCAGATGGCTTTCCGGAAGATGAGATTCATATATCTGCC 115  
Db 90030 GGATTTTGTCTTCAGATGGCTTTCCGGAAGATGAGATTCATATATCTGCC 90084

## RESULT 7

## HS21C080

LOCUS HS21C080 340000 bp DNA linear PRI 24-MAY-2000  
DEFINITION Homo sapiens chromosome 21 segment HS21C080.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## AUTHORS

## source

```
1..340000
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="21"
    /map="21q22.3"
    /clone="P70124"
    /clone_lib="RPC11,3-5 PAC library"
    /note="Accession No. AF064860"
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="21"
    /map="21q22.3"
    /clone="P206A10, 3' partial"
    /clone_lib="RPC11,3-5 PAC library"
    /note="Accession No. AF121782"
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="21"
    /map="21q22.3"
```

## Yaspo,M.L.

## Direct Submission

Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing  
Consortium: \* RIKEN Genomic Sciences Center, Human Genome Research  
Group \* Institute of Molecular Biotechnology, Genome Analysis \*  
Keio University School of Medicine, Dept. of Molecular Biology \*  
GBF, Dept. of Genome Analysis \* Max-Planck Institute for Molecular  
Genetics (addresses see below)  
The Chromosome 21 Mapping and Sequencing Consortium consists of  
\* RIKEN Genomic Sciences Center, Human Genome Research Group, \*  
Sagamihara 228-8555, Japan,  
\* e.mail: sakaki@gsc.riken.go.jp  
\* URL: http://hgp.gsc.riken.go.jp/  
and

\* Institute of Molecular Biotechnology, Genome Analysis, \*  
Beutenbergstrasse 11, D-07745 Jena, Germany,  
\* e.mail: gscj-submit@genome.imb-jena.de  
\* URL: http://genome.imb-jena.de/  
and

\* Keio University School of Medicine, Dept. of Molecular Biology, \*  
Tokyo 160-8582, Japan,  
\* e.mail: shimizu@mb.med.keio.ac.jp  
\* URL: http://adenine.dmb.med.keio.ac.jp/  
and

\* GBF, Dept. of Genome Analysis,  
\* Mascheroder Weg 1, D-38124 Braunschweig, Germany, \* e.mail:  
info.genome@gbf.de  
\* URL: http://genome.gbf.de/  
and

\* Max-Planck Institute for Molecular Genetics,  
\* Ihnestrasse 73, D-14195 Berlin, Germany,  
\* e.mail: info-chr21@molgen.mpg.de  
\* URL: http://chr21.rz-berlin.mpg.de/.

## Location/Qualifiers

## 1..340000

```
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
<1..125946
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="21"
    /map="21q22.3"
    /clone="P160L9, 5' partial"
    /clone_lib="RPC11,3-5 PAC library"
    /note="Accession No. AF121897"
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="21"
    /map="21q22.3"
    /clone="P70124"
    /clone_lib="RPC11,3-5 PAC library"
    /note="Accession No. AF064860"
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="21"
    /map="21q22.3"
    /clone="P206A10, 3' partial"
    /clone_lib="RPC11,3-5 PAC library"
    /note="Accession No. AF121782"
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="21"
    /map="21q22.3"
```

## FEATURES

## source

## source

## source

## source



```

/clone="BAC-291B3, 3' partial"
/clone_lib="BAC library"
/note="Accession No. AF064857"
repeat_region
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
repeat_region
2398..2495
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region
complement(2496..2878)
/note="THE1C"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region
2879..2931
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region
2932..3004
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region
3005..3074
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
gene
3473..19157
/gene="SH3BGR"
mRNA
join(<3473..3565,12649..12678,15331..15460,18739..19157)
/gene="SH3BGR"
CDS
join(3473..3565,12649..12678,15331..15426)
/gene="SH3BGR"
/note="Accession No. X93498"
/codon_start=1
/product="21-Glutamic Acid Rich protein 21-GARP"
/protein_id="CAB90445.1"
/db_xref="GI:7717370"
/db_xref="GOA:P55822"
/db_xref="UniProt/Swiss-Prot:P55822"
/translation="GSEKAEKGTEAQKGEVDGNLPEAQKNEREGETATSETE
3473..3565
/gene="SH3BGR"
/number=4
repeat_region
5542..5829
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region
complement(6887..7067)
/note="L1MD3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
repeat_region
complement(8482..8762)
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region
complement(8865..8964)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region
9131..9424
/note="AluSq"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region
complement(9723..9882)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region
complement(9989..10077)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED

```

---

```

complement(10078..10429)
/note="THE1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region
complement(10430..10651)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region
10980..11000
/note="(TTTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region
complement(11002..11282)
/note="AluSp"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region
complement(11315..11556)
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region
11702..11753
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
exon
12649..12678
/gene="SH3BGR"
/number=5
repeat_region
complement(13115..13425)
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region
complement(13598..13892)

```

---

```

Query Match      98.6%; Score 113.4; DB 9; Length 340000;
Best Local Similarity 99.1%; Pred. No. 2.2e-24;
Matches 114; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1  CCACCTCAGCCTCCTAGCATATAAACTAGACACATCCTCATGCTTTTGAGGTCCTAATCATTT 60
Db      163378 CCACCTCAGCCTCCTAGCATATAAACTAGACACATCCTCATGCTTTTGAGGTCCTAATCATTT 60
Qy      61  GGATTTTGTTCCTTTCAGATGGCTTTCCCGAAGATGAGATGATGATATATCTGCC 115
Db      163438 GGATTTTGTTCCTTTCAGATGGCTTTCCCGAAGATGAGATGATGATATATCTGCC 163492

RESULT 8
AB041414
LOCUS      AB041414
DEFINITION Pan troglodytes beta1.3-Galt 5 gene for UDP-Gal:GlcNAc
            beta1.3-galactosyltransferase 5, partial cds.
ACCESSION AB041414
VERSION   AB041414.1 GI:7593022
KEYWORDS
SOURCE    Pan troglodytes (chimpanzee)
ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE  1 (bases 1 to 1566)
            Liu, Y. and Saitou, N.
            Silver Project
            Published Only in Database (2000)
REFERENCE  2 (bases 1 to 1566)
            Liu, Y. and Saitou, N.
            Direct Submission
            Submitted (11-APR-2000) Naruya Saitou, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima,
            Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp,
            URL:http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790,
            Fax:81-559-81-6789)
FEATURES
            source
            1..1566
            /organism="Pan troglodytes"

```

```

/mol_type="genomic DNA"
/isoate="chimp-202"
/db_xref="taxon:9598"
/notes="human sequence used for primer design based on Acc#
AB020337"
<1..28
/number=3
29..675
/number=3
675..1566
/gene="beta1,3-Galt 5"
675..>1566
/gene="beta1,3-Galt 5"
/codon_start=1
/product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5"
/protein_id="BAA94499.1"
/db_xref="GI:7593023"
/translation="MAFFPKRRLMYVCLLVGALCVYFSMYSLNLFKEQSFVVKQGNF
LKLPTQDCQTPFLVLTSSHRQLAERMAIRQTGWKERTVKGKQLKTFLLIGTTSS
AAETKEVDQDSQGHGDIIOKFLDYVYNTLTMTMGIEWHFRFCPOAATVMTKDSMF
INVDYTELLKNNRTTRFTFTGLKNEFPIROPFSKFWKSEYPMWDRPPFCSTGT
VYFSGDVAQSVNVSVPYIKLEDVYVGLCLERLNIRLEELHSQTFPPGGLRFVSC
RFRRIVACHFIKPTLLDYNQ"
675..>1566
/gene="beta1,3-Galt 5"
/number=4

ORIGIN
Query Match          97.2%;   Score 111.8;   DB 9;   Length 1566;
Best local Similarity 98.3%;   Pred. No. 5.6e-24;
Matches 113;   Conservative 0;   Mismatches 2;   Indels 0;   Gaps 0;

QY      1   CCACCTCAGCCTCTAGCATAAACATAGACACATCCTCATGCTTTTGAGGTCTAAATCATT 60
      |||
Db      597 CCACCTCAGCCTCTAGCATAAACATAGACACATCCTCATGCTTTTGAGGTCTAAATCATT 656

QY      61   GGATTTTGTTCCCTTCAGATGGCTTTCCCGAAGATGAGATTGATGATATCTGCC 115
      |||
Db      657 GGATTTTGTTCCCTTCAGATGGCTTTCCCGAAGATGAGATTGATGATGTTTGCC 711

RESULT 9
AB041417
LOCUS
DEFINITION
Pongo pygmaeus beta1,3-Galt 5 gene for UDP-Gal:GlcNAc
beta1,3-galactosyltransferase 5, partial cds.
ACCESSION
AB041417
VERSION
AB041417.1 GI:7593028
KEYWORDS
SOURCE
Pongo pygmaeus (orangutan)
ORGANISM
Pongo pygmaeus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE
1 (bases 1 to 1579)
AUTHORS
Liu,Y. and Saitou,N.
TITLE
Silver Project
JOURNAL
Published Only in DataBase (2000)
REFERENCE
2 (bases 1 to 1579)
AUTHORS
Liu,Y. and Saitou,N.
TITLE
Direct Submission
JOURNAL
Submitted (11-APR-2000) Naruya Saitou, National Institute of
Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima,
Shizuoka 411-8540, Japan (E-mail:nsaitou@genetics.nig.ac.jp,
URL:http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790,
Fax:81-559-81-6789)
FEATURES
Location/Qualifiers
1..1579
/organism="Pongo pygmaeus"
/mol_type="genomic DNA"
/isoate="oran-Pol7"
/db_xref="taxon:9600"
/notes="human sequence used for primer design based on Acc#
AB020337"

```

```

source
1. .1570
/organism="Pan paniscus"
/mol_type="genomic DNA"
/isolate="bonobo-05"
/db_xref="taxon:9597"
/note="human sequence used for primer design based on Acc#
AB020337"
<1. .20
/number=3
21. .667
/number=3
668. .1570
/genes="beta1,3-GalT 5"
688. >.1570
/genes="beta1,3-GalT 5"
/codon_start=1
/product="UDP-Gal:GLCNAC beta1,3-galactosyltransferase 5"
/db_xref="GI:7593025"
/translation="WAFPMRLMYCLLVGLGALCYLFSYSLNLFEQSFVYKQGNF
LALPTDQCRQFPFVLVLVTSHKQAERMAIRQWTKERTVKGQKLTFFLLGTTSF
KLEKTEVDQESQRHGDIIQKDFLDGYNLTLKMMGIEWHVFQQAAYFMKTDSTG
INVDVTELLKKNTTRFFQGLKLNBPPIQPFSEKWFSEYKSPWDRYPFPCSTG
YVFGDVAOVYNSVESVPYIKLEDVFGVGLCLERLNIRLELHSQPTFFPGGLRFSVC
RFRRIVACHFIKPTLLIDYQALEN"
668. >.1570
/genes="beta1,3-GalT 5"
/number=4

ORIGIN
Query Match          95.8%;   Score 110.2;   DB 9;   Length 1570;
Best Local Similarity 97.4%;   Pred. No. 1.7e-23;
Matches 112;   Conservative 0;   Mismatches 3;   Indels 0;   Gaps 0;

QY      1   CCACCTCAGCCTCCTAGCATATAAACTAGACACATCCTCATGCTTTTGAGTGCTAAATCATT 60
Db      590 CCACCTCAACCTCCCTAGCATATAAACTAGACACATCCTCATGCTTTTGAGTGCTAAATCATT 649

QY      61   GGATTTTGTCTCTTCAGATGCGCTTTCCCGAAGATGAGATTCATGTATATCTGCC 115
Db      650 GGATTTTGTCTCTTCAGATGCGCTTTCCCGAAGATGAGATTCATGTATCTTTGCC 704

RESULT 12
AC150794
LOCUS
DEFINITION
AC150794
AC150794.3   GI:52219221
HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
Bos taurus (cow)
Bos taurus
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
1 (bases 1 to 170108)
Muzny, D., Adams, C., Agbai II, O., Allen, C., Alebrooks, S., Archer, P.,
Arredondo, H., Bandaranaike, D., Bangura, L., Beltran, B., Beltran, R.,
Beraducci, A., Biswal, K., Blyth, P., Bonham, H., Buhay, C., Burch, P.,
Cadoree, I., Canada, A., Cardenas, V., Carter, K., Cavazos, I.,
Chacko, J., Chahrouh, M., Chavez, D., Chen, A., Chen, G., Chen, R.,
Cheng, M.-T., Chu, J., Clerc, K., Cockrell, R., Coyle, M., Cree, A.,
Curry, S., Dai, W., Davila, M.L., Davis, C., Davy-Carroll, L., De
Anda, C., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H.,
Donlin, J., McCauley, S., Dugan-Rocha, S., Dunn, A., Durbin, K.,
Dziuda, D., Egan, A., Escotto, M., Espinosa, V., Eugene, C., Fa, M.,
Fernandez, S., Fernando, P., Flagg, N., Forbes, L., Foster, P.,
Fowler, G., Fu, Q., Fuh, E., Garcia, A., Garcia, R., Garner, T.,
Gaskin, C., Gench, S., Ghose, S., Gill, R., Gonzalez, D.,
Gonzalez-Garay, M., Guevara, W., Holder, M., Haaland, W., Haeberlen, K.,
Hall, B., Hamid, H., Hamilton, K., Harbes, B., Harris, R., Havlak, P.,
Hawes, A., Hawkins, E., Hayes, S., Hemphill, L., Hernandez, J.,

```

Hines, S., Hitchens, M., Hodgson, A., Hoques, M., Hollins, B., Howell, L. T., Hulyk, S., Hume, J., Imo, K., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Kalafus, K., Kelly, S., Keys, T., Khan, Z., King, L., Kovar, C., Kowis, A., Kowis, C., Lara, F., Leal, S., Lee, K., Lee, S., LeGall, F. I., Lemon, S., Lewis, L., Li, B., Li, Y., Li, Z., Linnell, M., Liu, W., Liu, Y.-S., Liu, Y., Liyanage, D., London, P., Lopez, J., Lorensuhewa, L., Lozado, R., Luk, T., Madu, R., Maheshwari, M., Mahoney, C., Malloy, K., Mansouri, D., Martinez, E., McClelland, H., McPherson, J., Mercadao, C., Metsker, M., Milosavljevic, A., Minja, E., Morgan, M., Morris, S., Munidasa, M., Murray, D., Nazareth, L., Ngo, D., Nguyen, N., Norwig-Eastaugh, E., Nott, A., Nwaokelemeh, O., Oregon, M., Ochi-Okorie, C., Odeh, E., Okwuonu, G., Okwuonu, K., Parker, D., Pasternak, S., Patel, B., Patel, V., Paul, H., Perez, A., Perez, L., Petrosino, J., Pham, T., Primus, E., Pu, L.-L., Puazo, M., Qin, X., Quinn, A., Quiroz, J., Rabata, D., Rachlin, E., Reigh, R., Ren, Y., Reuter, M., Richards, S., Rives, C., Rodriguez, F., Rojas, A., Ruiz, S. J., Sana, M., Sanders, W., Santibanez, J., Santos, R., Savary, G., Scherer, S., Shen, H., Shen, Y., Sisson, I., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Svatek, A., Taylor, E., Taylor, T., Thomas, N., Thorn, R., Thornton, R., Trejos, Z., Umani, K., Vargo, C., Verduzco, D., Villasana, D., Virk, D., Volkov, A., Waldron, B., Walker, B., Wang, Q., Wang, S., Warren, J., Wei, X., Wheeler, D., Williams, G., Williams, R., Worley, K., Wright, R., Wu, J., Yakub, S., Yan, K., Yaun, Y., Yu, F., Zhang, J., Zhang, L., Zhang, Z., Zhou, J., Weinstock, G. and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 170108)  
Worley, K.C.  
Direct Submission  
Submitted (10-AUG-2004) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 170108)  
Worley, K.C.  
Direct Submission  
Submitted (17-SEP-2004) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Sep 17, 2004 this sequence version replaced gi:51172637.  
-----  
Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help.tmc.edu  
-----  
Project Information  
Center project name: FEXJ  
Center clone name: CH240-397A7  
-----  
Summary Statistics  
Sequencing vector: Plasmid  
Chemistry: Dye-terminator Big Dye: 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 163512 bases at least Q40  
Consensus quality: 163737 bases at least Q30  
Consensus quality: 163990 bases at least Q20  
Estimated insert size: 167831; sum-of-contigs estimation  
Quality coverage: 11x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))  
\* The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. The indicated order and orientation of each sequence has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing.  
\* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
\* This sequence will be replaced

\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

1 16110: contig of 16110 bp in length  
16111 16760: gap of 650 bp  
16761 69838: contig of 53078 bp in length  
69839 71638: gap of 1800 bp  
71639 108013: contig of 36375 bp in length  
108014 109713: gap of 1700 bp  
109714 111775: contig of 2062 bp in length  
111776 113375: gap of 1600 bp  
113376 157026: contig of 43651 bp in length  
157027 158026: gap of 1000 bp  
158027 170108: contig of 12082 bp in length.

FEATURES  
Location/Qualifiers  
1..170108  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9913"  
/clone="CH240-397A7"  
1..16110  
/note="assembly\_name:Contig32"  
16111..16760  
/note="assembly\_name:gap"  
16761..69838  
/note="assembly\_name:Contig34"  
69839..71638  
/note="assembly\_name:gap"  
71639..108013  
/note="assembly\_name:Contig33"  
108014..109713  
/note="assembly\_name:gap"  
109714..111775  
/note="assembly\_name:Contig31"  
111776..113375  
/note="assembly\_name:gap"  
113376..157026  
/note="assembly\_name:Contig36"  
157027..158026  
/note="assembly\_name:gap"  
158027..170108  
/note="assembly\_name:Contig37"

ORIGIN  
Query Match 50.6%; Score 58.2; DB 2; Length 170108;  
Best Local Similarity 75.7%; Pred. No. 2.7e-07;  
Matches 87; Conservative 0; Mismatches 23; Indels 5; Gaps 1;  
Qy 1 CCACCTCAGCCTCCTAGCATAAACATGACACATCCTCATGCTTTTGGAGGTCTATCATTT 60  
Db 152769 CCATCTCAGCCACCAGGCGCAAGCTATCTCTCTCT-----TTTGGAGGTCTATCATTT 152823  
Qy 61 GGATTTTGTCTTCTTCAGATGGCTTTCCGAGATGAGATGATGATGATATATCTGCC 115  
Db 152824 TGAATGTTTCTTCTTCAGATGGCTTACGTGAAGATGAGATGGGTATATATTTCCC 152878

RESULT 13  
AC126579/c  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-210M15, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*  
AC126579  
AC126579.4 GI:25138868  
VERSION HTG; HTGS PHASE2; HTGS DRAFT; HTGS\_ENRICHED.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 243047)  
AUTHORS Munz, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GZJB
Center clone name: CH230-210M15
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 198187 bases at least Q40
Consensus quality: 201579 bases at least Q30
Consensus quality: 204254 bases at least Q20
Estimated insert size: 209786; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  consists of 1 contigs. Gaps between the contigs
  are represented as runs of N. The order of the pieces
  is believed to be correct as given, however the sizes
  of the gaps between them are based on estimates that have
  been provided by the submitter.
* This sequence will be replaced
  by the finished sequence as soon as it is available and
  the accession number will be preserved.
  1 243047: contig of 243047 bp in length.
      Location/Qualifiers
        1. 243047
          /organism="Rattus norvegicus"
          /mol_type="genomic DNA"
          /db_xref="taxon:10116"
          /clone="CH230-210M15"
        1. 1635
          /note="wgs_contig"
signature
  32.3%; Score 37.2; DB 2; Length 243047;
  Similarity 65.9%; Pred. No. 0.89;
  54; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
  33 ATCCCTACGCTTTTGAGCTCTAATCATTTGGATTGTTCCTTTTCAGATGGCTTTCCGAA 92
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  57 ATTTTCAGTATTAAAGAGATGATCATGTGATTTTTTTTCTTTCTACTTTGTGTGGCA 1071

```

HSA6078                      933 bp                      DNA                      linear                      PRI 11-MAY-2000  
 Homo sapiens beta3gal-T6 gene.  
 AJ006078

Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 Amado M., Carneiro, F. and Clausen, H.  
 Cloning and expression of two beta-1,3-galactosyltransferases:  
 beta3gal-T5 and beta3gal-T6  
 Unpublished  
 2 (bases 1 to 933)  
 Amado, M.  
 Direct Submission  
 Submitted (11-MAY-1998) Amado M., Department of Oral Diagnostics,  
 Royal dental School, Norre Alle 20, 2200 Copenhagen, DENMARK  
 Location/Qualifiers



```
misc_feature (NID:g13906660)"
241. .819 /note="similar to Homo sapiens EST BM006922
(NID:g16521276)"
misc_feature 241. .813 /note="similar to Homo sapiens EST BE619661
(NID:g9890599)"
misc_feature 241. .699 /note="similar to Homo sapiens EST BI856910
(NID:g15997657)"
misc_feature 241. .691 /note="similar to Homo sapiens EST BE173595
(NID:g8636321)"
misc_feature 241. .392 /note="similar to Homo sapiens EST BI868897
(NID:g16042570)"
misc_feature 241. .392 /note="similar to Homo sapiens EST BM019635
(NID:g16533989)"
misc_feature 241. .361 /note="similar to Homo sapiens EST AI042168 (NID:g3281362)
cy37b10.x1"
misc_feature 244. .361 /note="similar to Homo sapiens EST BG708944
(NID:g13986789)"
misc_feature 267. .369 /note="similar to Homo sapiens EST BE778176
(NID:g10199374)"
misc_feature 274. .819 /note="similar to Homo sapiens EST BE561211
(NID:g98049311)"
misc_feature 285. .819 /note="similar to Homo sapiens EST BG341756
(NID:g131148194)"
misc_feature 289. .819 /note="similar to Homo sapiens EST BM010255
(NID:g16524609)"
misc_feature 414. .923 /note="similar to EST BM171946 (NID:g17311509)"
misc_feature 446. .923 /note="similar to Bos taurus EST AV597915 (NID:g9715406)"
misc_feature 453. .819 /note="similar to Homo sapiens EST AW175759
(NID:g6441705)"
misc_feature 453. .819 /note="similar to Homo sapiens EST AW601923
(NID:g7306662)"
misc_feature 453. .819 /note="similar to Homo sapiens EST BG708944
(NID:g13986789)"
misc_feature 453. .814 /note="similar to Homo sapiens EST BE778176
(NID:g10199374)"
misc_feature 453. .814 /note="similar to Homo sapiens EST BM019635
(NID:g16533989)"
misc_feature 453. .797 /note="similar to Homo sapiens EST BE535420
(NID:g9764065)"
misc_feature 453. .485 /note="similar to Homo sapiens EST BE905008
(NID:g10397852)"
misc_feature 464. .819 /note="similar to Homo sapiens EST BI197251
(NID:g14652271)"
misc_feature 464. .804 /note="similar to Homo sapiens EST AI042168 (NID:g3281362)
cy37b10.x1"
misc_feature 464. .485 /note="similar to Homo sapiens EST BI868897
(NID:g16042570)"
misc_feature 541. .814 /note="similar to Homo sapiens EST BI868897
```

```
misc_feature (NID:g16042570)"
541. .756 /note="similar to Homo sapiens EST BE905008
(NID:g10397852)"
misc_feature 1179. 1202 /note="similar to Homo sapiens EST BG708944
(NID:g13986789)"
repeat_region 2333. .2585 /rpt_family="Alu"
repeat_region 3106. .3655 /rpt_family="MaLR"
repeat_region 3812. .3840 /rpt_family="AT-rich"
repeat_region 4393. .4702

Query Match 32.0%; Score 36.8; DB 9; Length 131395;
Best Local Similarity 61.5%; Pred. No. 1.2;
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 20 TAAAACTAGACACATCCTCATGCTTTTGAGGTCTAAATCATTGGATTTTGTCCTTTCAGA 79
Db 28052 TAAATATTAGTCACATGGTCTTTTCATTGCAGGATGAATTTATTGAATGTGCTGGCATGATCA 28111

Qy 80 TGGCTTTTCCCGAAGATGAGATTGATGTATATCTGCC 115
Db 28112 TTGCTAAACAGAAACTCAGAAATGGCCTACATCTGCC 28147

Search completed: September 22, 2005, 14:50:04
Job time : 529.906 secs
```

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2005, 05:52:47 ; Search time 68.5965 Seconds  
(without alignments)  
9924.272 Million cell updates/sec

Title: US-10-777-828-8\_COPY\_1\_115

Perfect score: 115

Sequence: 1 ccacctcagctcctagcat.....gagattgatgtatctgtcc 115

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04.\*

- 1: Geneseq1980s.\*
- 2: Geneseq1990s.\*
- 3: Geneseq2000s.\*
- 4: Geneseq2001as.\*
- 5: Geneseq2001bs.\*
- 6: Geneseq2002as.\*
- 7: Geneseq2002bs.\*
- 8: Geneseq2003as.\*
- 9: Geneseq2003bs.\*
- 10: Geneseq2003cs.\*
- 11: Geneseq2003ds.\*
- 12: Geneseq2004as.\*
- 13: Geneseq2004bs.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115	100.0	1011	3	Aaa27959 Human bet
2	113.4	98.6	596	12	Ach79926 Human gen
3	113.4	98.6	3409	10	Abz56919 B3GALT nu
4	113.4	98.6	10562	3	Aaa93876 Human bet
5	37	32.2	933	10	Abz56916 B3GALT nu
6	35.8	31.1	2629	10	Abz56918 B3GALT nu
7	35.8	31.1	2762	10	Abz56917 B3GALT nu
8	35.8	31.1	2775	3	Aaa93875 Human bet
9	35.4	30.8	933	12	Adk68456 Human bet
10	33	28.7	78025	8	Abq77404 Human SEL
11	32	27.8	110000	6	Continuation (14 o
12	32	27.8	110000	6	Continuation (4 of
13	30.8	26.8	3796	6	Abq67128 Human ang
14	30.6	26.6	2000	6	Abz16396 Arabidops
15	30.4	26.4	2000	6	Abz15338 Arabidops
16	30.4	26.4	335199	10	Adc24703 Human wil
17	30.2	26.3	260027	11	Acn44046 Human gen
18	30	26.1	99960	3	Aaz50905 Human TBC
19	29.8	25.9	502	4	Aah11443 Human cDN
20	29.8	25.9	682	4	Aah33538 Human col

21	29.8	25.9	769	2	AAZ17514	Aaz17514 Human gen
22	29.8	25.9	1578	4	AAS33072	Aas33072 DNA encod
23	29.8	25.9	1817	6	ABQ60886	Abq60886 FLJ20206
24	29.8	25.9	2971	4	AAH16199	Aah16199 Human cDN
25	29.8	25.9	3657	12	ADK00670	Adk00670 HOMO prot
26	29.8	25.9	4506	6	ABS58378	Abs58378 Protein m
27	29.4	25.6	1286	4	AAS06098	Aas06098 Human PRO
28	29.4	25.6	1364	10	ADB97571	Adb97571 Human MTS
29	29.4	25.6	1443	8	ABX56301	Abx56301 Human NOV
30	29.4	25.6	103747	6	ABQ88139	Abq88139 Human ost
31	29.2	25.4	619	6	AAC41190	Aac41190 Arabidops
32	29.2	25.4	5683	6	ABL33807	Ab133807 Human imm
33	29.2	25.4	5683	6	ABL34591	Ab134591 Human met
34	29.2	25.4	5683	6	ABL70400	Ab170400 Chemical
35	29.2	25.4	5683	7	ADS99852	AdS99852 Complemen
36	29.2	25.4	5683	13	ADS89332	AdS89332 Oligonucl
37	29.2	25.4	5683	10	ADBS4182	Adb54182 Pretreat
38	29.2	25.4	71843	8	ACC79619	Acc79619 Zebrafish
39	29	25.2	359	2	AAV77281	Aav77281 Staphyloc
40	29	25.2	996	3	AAC69440	Aac69440 Human sec
41	29	25.2	2000	8	ADA72146	Ada72146 Rice gene
42	28.8	25.0	936	12	ADN73186	Adn73186 Thale cre
43	28.8	25.0	1212	3	AAC47268	Aac47268 Arabidops
44	28.8	25.0	1213	3	AAC39838	Aac39838 Arabidops
45	28.8	25.0	2330	3	AAI38011	Aai38011 Diabete

ALIGNMENTS

RESULT 1						
AAA27959						
ID	AAA27959	standard; DNA; 1011 BP.				
AC	AAA27959;					
XX						
DT	15-AUG-2000	(first entry)				
XX						
DE	Human beta3Gal-T5	gene sequence.				
XX						
KW	UDP-D-galactose; beta-N-acetylglucosamine beta1,3-galactosyltransferase;					
KW	Beta3gal-T5; ss; human; chromosome 21q22.3; galactosylation;					
KW	beta1,3-galactosyl glycosylated saccharide production; glycopeptide;					
KW	glycoprotein.					
XX						
OS	Homo sapiens.					
XX						
FH	Key	Location/Qualifiers				
FT	CDS	79..1011				
FT		/*tag= a				
FT		/product= "Beta3Gal-T5"				
FT	primer_bind	complement(79..98)				
FT		/*tag= b				
FT	primer_bind	complement(150..170)				
FT		/*tag= c				
FT	primer_bind	991..1011				
FT		/*tag= d				
PN	WO200029558-A1.					
XX						
PD	25-MAY-2000.					
XX						
PF	11-NOV-1999;	99WO-US026807.				
XX						
PR	13-NOV-1998;	98DK-00001483.				
XX						
PA	(CLAU/)	CLAUSEN H.				
XX						
PI	Clausen H, Amado M;					
XX						
DR	WPI; 2000-399728/34.					
DR	P-ESDB; AAY94641.					
XX						

PT Novel nucleic acid sequence encoding human UDP-galactose:beta-N-  
PT acetylglucosamine beta1,3-galactosyltransferase useful for obtaining beta  
PT 1,3-galactosyl glycosylated saccharides and glycopeptides or  
PT glycoproteins.

Claim 7; Fig 1; 74pp; English.

XX The present invention relates to a nucleic acid sequence encoding UDP-D-  
XX galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase  
XX (beta3Gal-T5). Beta3 transferases add galactose to the hydroxy group at  
XX carbon 3 of 2-acetamido-2-deoxy-D-glucose (GlcNAc). The present sequence  
XX represents the human beta3Gal-T5 gene sequence. The beta3Gal-T5 gene is  
XX located on human chromosome 21q22.3. Beta3Gal-T5 is a type II  
XX transmembrane glycoprotein. The invention also relates to the beta3Gal-T5  
XX protein sequence, a nucleic acid vector comprising the beta3Gal-T5  
XX nucleotide sequence, a host cell comprising the vector, and a method for  
XX the production of the beta3Gal-T5 protein from the host cells. The  
XX methods of the invention can be used for recombinant production of  
XX beta3Gal-T5 for use as a catalyst and for recombinant production of  
XX peptides or proteins with appropriate galactosylation. The beta3Gal-T5  
XX protein can be used to obtain beta1,3-galactosyl glycosylated  
XX saccharides, glycopeptides or glycoproteins

Sequence 1011 BP; 247 A; 256 C; 251 G; 257 T; 0 U; 0 Other;

Query Match 100.0%; Score 115; DB 3; Length 1011;  
Best Local Similarity 100.0%; Pred. No. 6e-28;  
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACCTCAGCCTCCTAGCATAAACTAGACATCCTCATGCTTTGAGGCTTAATCATT 60  
Db 1 CCACCTCAGCCTCCTAGCATAAACTAGACATCCTCATGCTTTGAGGCTTAATCATT 60

Qy 61 GGATTTGTTCTTCAGATGGCTTTCCGAGATGAGATTGATATATCTGCC 115  
Db 61 GGATTTGTTCTTCAGATGGCTTTCCGAGATGAGATTGATATATCTGCC 115

RESULT 2  
ACH79926  
ID ACH79926 standard; DNA; 596 BP.

XX ACH79926;

XX 29-JUL-2004 (first entry)

XX Human genome derived single exon probe #13121.

XX Human; probe: ss; gene expression; single exon probe; microarray;  
XX alternative splicing event; genomic alteration.

XX Homo sapiens.

XX US2003194704-A1.

XX 16-OCT-2003.

XX 03-APR-2002; 2002US-00029386.

XX 03-APR-2002; 2002US-00029386.

XX (PENN/) PENN S G.  
XX (RANK/) RANK D R.  
XX (HANK/) HANKEL D K.

XX Penn SG, Rank DR, Hanzel DK;

XX WPI; 2004-119264/12.

XX New human genome-derived single exon nucleic acid probes useful for human  
XX gene expression analysis, for identifying or characterizing alternative  
XX splicing events, for assessing genomic alterations or as tools for  
XX surveying tissues.

XX Claim 15; SEQ ID NO 13121; 80pp; English.

XX The invention relates to a nucleic acid probe for measuring human gene  
XX expression, comprising any of the 27,400 fully defined nucleotide  
XX sequences in the specification, or their complements or fragments, and  
XX encoding at least 8 amino acids of any of the 6888 amino acid sequences  
XX fully defined in the specification. The probe is a single exon probe that  
XX hybridises under high stringency conditions to a nucleic acid molecule  
XX expressed in human cells or tissues. Also included are a spatially-  
XX addressable set of single exon nucleic acid probes for measuring human  
XX gene expression (comprising a plurality of single exon nucleic acid  
XX probes cited above, where each of the plurality of probes is separately  
XX and addressably isolatable or amplifiable from the plurality), a single  
XX exon microarray for measuring human gene expression, a method of  
XX measuring human gene expression, a vector comprising the single exon  
XX probe cited above, an ORF-encoded peptide comprising at least 8  
XX contiguous amino acids of any of the above-mentioned amino acid  
XX sequences (optionally with conservative amino acid substitutions), an  
XX isolated antibody that binds specifically to a peptide cited above,  
XX methods of selling and/or licensing single exon probes or microarrays to  
XX a customer desiring to measure gene expression, and a method of providing  
XX human gene expression data by subscription, and a computer-readable  
XX storage medium which contains a database having a plurality of records  
XX (each record including data on the expression of a single exon probe  
XX cited above. The probe, methods and apparatus are useful in gene  
XX expression analysis. The probes may be used as tools for surveying  
XX tissues to detect the presence of expressed messages that contain their  
XX specific exon, or in constructing genome-derived single exon microarrays.  
XX In addition, the probes are used in identifying and characterizing  
XX alternative splicing events, in detecting and characterising gross  
XX alterations in the genomic locus that includes their exon, in assessing  
XX smaller genomic alterations, in priming the synthesis of nucleic acids,  
XX or in expressing the ORF-encoded peptide. The present sequence is a human  
XX single exon probe of the invention. Note: The sequence data for this  
XX patent did not form part of the printed specification, but was obtained  
XX in electronic format directly from USPTO at  
XX seqdata.uspto.gov/sequence.html?DocID=20030194704

XX Sequence 596 BP; 149 A; 150 C; 141 G; 156 T; 0 U; 0 Other;

Query Match 98.6%; Score 113.4; DB 12; Length 596;  
Best Local Similarity 99.1%; Pred. No. 1.7e-27;  
Matches 114; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCACCTCAGCCTCCTAGCATAAACTAGACATCCTCATGCTTTGAGGCTTAATCATT 60  
Db 187 CCACCTCAGCCTCCTAGCATAAACTAGACATCCTCATGCTTTGAGGCTTAATCATT 246

Qy 61 GGATTTGTTCTTCAGATGGCTTTCCGAGATGAGATTGATATATCTGCC 115  
Db 247 GGATTTGTTCTTCAGATGGCTTTCCGAGATGAGATTGATATATCTGCC 301

RESULT 3

ABZ56919

ID ABZ56919 standard; DNA; 3409 BP.

XX ABZ56919;

XX 04-APR-2003 (first entry)

XX B3GALT nucleic acid sequence # SEQ ID 11.

XX B3GALT; beta-1,3 galactosyltransferase; p53; cancer; breast; colon;  
XX kidney; lung; ovary; gene; ds.  
XX Homo sapiens.

XX WO200299044-A2.

XX 12-DEC-2002.

PF 02-JUN-2002; 2002WO-US017356.  
XX  
PR 05-JUN-2001; 2001US-0296076P.  
PR 10-OCT-2001; 2001US-0328605P.  
PR 15-FEB-2002; 2002US-0357253P.  
XX  
PA (EXEL-) EXELIXIS INC.  
XX  
PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
XX WPI; 2003-156849/15.  
XX  
DR Identifying p53 pathway modulating agents with B3GALT genes, useful for  
XX the diagnosis and treatment of disorders associated with defects in the  
XX p53 pathway, such as cancer of the breast, colon, kidneys, lung and  
XX ovary.  
XX  
PS Disclosure; Page 55-57; 82pp; English.  
XX  
XX The invention relates to identifying a candidate p53 pathway modulating  
XX agent in humans that is referred to in the specification as BGALT (beta-  
XX 1,3 galactosyltransferase). Also disclosed is a method for diagnosing a  
XX disease in a patient, by contacting a sample with a probe for B3GALT  
XX expression, and comparing the results with a control, and determining  
XX whether the results indicate a likelihood of disease. Methods and  
XX compositions of the invention are useful for the diagnosis and treatment  
XX of disorders associated with defects in the p53 pathway, such as cancer  
XX of the breast, colon, kidneys, lung and ovary. The current sequence  
XX represents a B3GALT nucleic acid sequence referred to in the disclosure  
XX of the invention  
XX  
SQ Sequence 3409 BP; 833 A; 842 C; 833 G; 901 T; 0 U; 0 Other;  
Query Match 98.6%; Score 113.4; DB 10; Length 3409;  
Best Local Similarity 99.1%; Pred. No. 2.9e-27;  
Matches 114; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CCACCTCAGCCTCCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGTCTAATCATT 60  
Db 1003 CCACCTCAGCCTCCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGTCTAATCATT 1062  
QY 61 GGATTTTGTCTTTTCAGATGCGTTTCCGAGATGAGATTGATATATCTGCC 115  
Db 1063 GGATTTTGTCTTTTCAGATGCGTTTCCGAGATGAGATTGATATATCTGCC 1117  
RESULT 4  
AAA93876  
ID AAA93876 standard; DNA; 10562 BP.  
AC AAA93876;  
XX  
XX 15-JAN-2001 (first entry)  
XX Human beta3Gal-T5 encoding DNA.  
XX  
XX Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;  
XX digestive system; beta3gal-T5; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200050608-A1.  
XX  
XX 31-AUG-2000.  
XX  
XX 24-FEB-2000; 2000WO-JP001070.  
XX  
XX 25-FEB-1999; 99JP-00047571.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
XX Narimatsu H, Ishihiki S, Togayachi A, Sasaki K;  
XX

DR WPI; 2000-549409/50.  
XX  
XX Beta-1,3 galactose transferase and DNA encoding it, useful for synthesis  
XX of type 1 sialyl Lewis, a carbohydrate for treatment of digestive system  
XX cancer.  
XX  
XX Claim 31; Page 103-111; 123pp; Japanese.  
XX  
XX This invention relates to a polypeptide (I) with beta-1,3 galactose  
XX transferase activity, or variants of (I) comprising amino acid additions,  
XX deletions and/or substitutions. Included in the invention is DNA encoding  
XX all or part of (I); expression vectors containing the DNA, host cells  
XX transformed by the vectors; a method for the preparation of the  
XX polypeptide by culture of the transformants or by expression in the milk  
XX of a transgenic mammal, and antibodies recognising (I). The Beta-1,3  
XX galactose transferase protein transfers galactose by beta-1,3 bonding to  
XX N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as  
XX GlcNAc-beta1-3gal-beta1-4Glc) to give Galbeta1-3GlcNAc. The protein and DNA  
XX encoding it are useful for the treatment and diagnosis of cancer of the  
XX digestive system. The present sequence represents a Beta3Gal-T5 encoding  
XX DNA sequence  
XX  
SQ Sequence 10562 BP; 2610 A; 2415 C; 2574 G; 2963 T; 0 U; 0 Other;  
Query Match 98.6%; Score 113.4; DB 3; Length 10562;  
Best Local Similarity 99.1%; Pred. No. 4.1e-27;  
Matches 114; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CCACCTCAGCCTCCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGTCTAATCATT 60  
Db 8156 CCACCTCAGCCTCCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGTCTAATCATT 8215  
QY 61 GGATTTTGTCTTTTCAGATGCGTTTCCGAGATGAGATTGATATATCTGCC 115  
Db 8216 GGATTTTGTCTTTTCAGATGCGTTTCCGAGATGAGATTGATATATCTGCC 8270  
RESULT 5  
ABZ56916  
ID ABZ56916 standard; DNA; 933 BP.  
AC ABZ56916;  
XX  
XX 04-APR-2003 (first entry)  
XX  
XX B3GALT nucleic acid sequence # SEQ ID 8.  
XX  
XX B3GALT; beta-1,3 galactosyltransferase; p53; cancer; breast; colon;  
XX kidney; lung; ovary; gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200299044-A2.  
XX  
XX 12-DEC-2002.  
XX  
XX 02-JUN-2002; 2002WO-US017356.  
XX  
XX 05-JUN-2001; 2001US-0296076P.  
XX  
XX 10-OCT-2001; 2001US-0328605P.  
XX  
XX 15-FEB-2002; 2002US-0357253P.  
XX  
XX (EXEL-) EXELIXIS INC.  
XX  
XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
XX WPI; 2003-156849/15.  
XX  
XX Identifying p53 pathway modulating agents with B3GALT genes, useful for  
XX the diagnosis and treatment of disorders associated with defects in the  
XX p53 pathway, such as cancer of the breast, colon, kidneys, lung and  
XX ovary.  
XX

PS Disclosure; Page 51-52; 82pp; English.

XX The invention relates to identifying a candidate p53 pathway modulating  
CC agent in humans that is referred to in the specification as BGALT (beta-  
CC 1,3 galactosyltransferase). Also disclosed is a method for diagnosing a  
CC disease in a patient, by contacting a sample with a probe for B3GALT  
CC expression, and comparing the results with a control, and determining  
CC whether the results indicate a likelihood of disease. Methods and  
CC compositions of the invention are useful for the diagnosis and treatment  
CC of disorders associated with defects in the p53 pathway, such as cancer  
CC of the breast, colon, kidneys, lung and ovary. The current sequence  
CC represents a B3GALT nucleic acid sequence referred to in the disclosure  
CC of the invention

XX Sequence 933 BP; 229 A; 234 C; 240 G; 229 T; 0 U; 1 Other;

Query Match 32.2%; Score 37; DB 10; Length 933;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 79 ATGGCTTTCCCGAGATGAGATGATGATATATCTGCCC 115  
Db 1 ATGGCTTTCCCGAGATGAGATGATGATATATCTGCCC 37

#### RESULT 6

ABZ56918  
ID ABZ56918 standard; DNA; 2629 BP.

AC ABZ56918;

DT 04-APR-2003 (first entry)

XX B3GALT nucleic acid sequence # SEQ ID 10.

XX B3GALT; beta-1,3 galactosyltransferase; p53; cancer; breast; colon;  
KW kidney; lung; ovary; gene; ds.

XX Homo sapiens.

XX WO200299044-A2.

XX 12-DEC-2002.

XX 02-JUN-2002; 2002WO-US017356.

XX 05-JUN-2001; 2001US-0296076P.

XX 10-OCT-2001; 2001US-0328605P.

XX 15-FEB-2002; 2002US-0357253P.

XX (EXEL-) EXELIXIS INC.

XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;

XX WPI; 2003-156849/15.

XX Identifying p53 pathway modulating agents with B3GALT genes, useful for  
PT the diagnosis and treatment of disorders associated with defects in the  
PT p53 pathway, such as cancer of the breast, colon, kidneys, lung and  
PT ovary.

PS Disclosure; Page 53-55; 82pp; English.

XX The invention relates to identifying a candidate p53 pathway modulating  
CC agent in humans that is referred to in the specification as BGALT (beta-  
CC 1,3 galactosyltransferase). Also disclosed is a method for diagnosing a  
CC disease in a patient, by contacting a sample with a probe for B3GALT  
CC expression, and comparing the results with a control, and determining  
CC whether the results indicate a likelihood of disease. Methods and  
CC compositions of the invention are useful for the diagnosis and treatment  
CC of disorders associated with defects in the p53 pathway, such as cancer  
CC of the breast, colon, kidneys, lung and ovary. The current sequence  
CC represents a B3GALT nucleic acid sequence referred to in the disclosure

CC of the invention

XX Sequence 2629 BP; 632 A; 665 C; 644 G; 688 T; 0 U; 0 Other;  
Query Match 31.1%; Score 35.8; DB 10; Length 2629;  
Best Local Similarity 94.9%; Pred. No. 0.16;  
Matches 37; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 77 AGATGGCTTTCCCGAGATGAGATGATGATATATCTGCCC 115  
Db 299 AAATGGCTTTCCCGAGATGAGATGATGATATATTTGCC 337

#### RESULT 7

ABZ56917  
ID ABZ56917 standard; DNA; 2762 BP.

AC ABZ56917;

DT 04-APR-2003 (first entry)

XX B3GALT nucleic acid sequence # SEQ ID 9.

XX B3GALT; beta-1,3 galactosyltransferase; p53; cancer; breast; colon;  
KW kidney; lung; ovary; gene; ds.

XX Homo sapiens.

XX WO200299044-A2.

XX 12-DEC-2002.

XX 02-JUN-2002; 2002WO-US017356.

XX 05-JUN-2001; 2001US-0296076P.

XX 10-OCT-2001; 2001US-0328605P.

XX 15-FEB-2002; 2002US-0357253P.

XX (EXEL-) EXELIXIS INC.

XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
XX WPI; 2003-156849/15.

XX Identifying p53 pathway modulating agents with B3GALT genes, useful for  
PT the diagnosis and treatment of disorders associated with defects in the  
PT p53 pathway, such as cancer of the breast, colon, kidneys, lung and  
PT ovary.

XX Example 5; Page 52-53; 82pp; English.

XX The invention relates to identifying a candidate p53 pathway modulating  
CC agent in humans that is referred to in the specification as BGALT (beta-  
CC 1,3 galactosyltransferase). Also disclosed is a method for diagnosing a  
CC disease in a patient, by contacting a sample with a probe for B3GALT  
CC expression, and comparing the results with a control, and determining  
CC whether the results indicate a likelihood of disease. Methods and  
CC compositions of the invention are useful for the diagnosis and treatment  
CC of disorders associated with defects in the p53 pathway, such as cancer  
CC of the breast, colon, kidneys, lung and ovary. The current sequence  
CC represents a B3GALT nucleic acid sequence referred to in an example from  
CC the invention

XX Sequence 2762 BP; 651 A; 703 C; 679 G; 729 T; 0 U; 0 Other;  
Query Match 31.1%; Score 35.8; DB 10; Length 2762;  
Best Local Similarity 94.9%; Pred. No. 0.16;  
Matches 37; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 77 AGATGGCTTTCCCGAGATGAGATGATGATATATCTGCCC 115  
Db 432 AAATGGCTTTCCCGAGATGAGATGATGATATATTTGCC 470

```

RESULT 8
ID AAA93875 standard; DNA; 2775 BP.
XX
AC AAA93875;
XX
DT 15-JAN-2001 (first entry)
XX
DE Human beta-1,3 galactose transferase encoding DNA.
XX
KW Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
KW digestive system; ds.
XX
OS Homo sapiens.
XX
PN WO200050608-A1.
XX
PD 31-AUG-2000.
XX
PF 24-FEB-2000; 2000WO-JP001070.
XX
PR 25-FEB-1999; 95JP-00047571.
XX
PA (KYOM ) KYOMA HAKKO KOGYO KK.
XX
PI Narimatsu H, Isshiki S, Togayachi A, Sasaki K;
XX
DR WPI; 2000-549409/50.
DR P-PSDB; AAB93875.
XX
PT Beta-1,3 galactose transferase and DNA encoding it, useful for synthesis
PT of type 1 sialyl Lewis, a carbohydrate for treatment of digestive system
PT cancer.
XX
PS Claim 5; Page 99-102; 123pp; Japanese.
XX
CC This invention relates to a polypeptide (I) with beta-1,3 galactose
CC transferase activity, or variants of (I) comprising amino acid additions,
CC deletions and/or substitutions. Included in the invention is DNA encoding
CC all or part of (I); expression vectors containing the DNA, host cells
CC transformed by the vectors; a method for the preparation of the
CC polypeptide by culture of the transformants or by expression in the milk
CC of a transgenic mammal, and antibodies recognising (I). The Beta-1,3
CC galactose transferase protein transfers galactose by beta-1,3 bonding to
CC N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as
CC GlcNAc) to give Galbeta1-3GlcNAc. The protein and DNA
CC encoding it are useful for the treatment and diagnosis of cancer of the
CC digestive system. The present sequence represents Beta-1,3 galactose
CC transferase encoding DNA
XX
SQ Sequence 2775 BP; 681 A; 698 C; 669 G; 727 T; 0 U; 0 Other;

Query Match 31.1%; Score 35.8; DB 3; Length 2775;
Best Local Similarity 94.9%; Pred. No. 0.16;
Matches 37; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 77 AGATGGCTTCCCGAAGATGAGATTGATGTATATCTGCC 115
DB 400 AAATGGCTTCCCGAAGATGAGATTGATGTATATCTGCC 438

RESULT 9
ADK68456
ID ADK68456 standard; cDNA; 933 BP.
XX
AC ADK68456;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human beta-1,3-galactosyl transferase cDNA SeqID 1.
XX
KW human; gene; ss; saccharide binding protein; maltose binding protein;

```

---

```

KW beta-1,3-galactosyl transferase; inflammation; infectious disease;
KW cancer metastasis suppression; dairy product; antiinflammatory;
KW antimicrobial; cytostatic.
OS Homo sapiens.
XX
Key Location/Qualifiers
FH 1.933
FT /*tag= a
FT /product= "Beta-1,3-galactosyl transferase protein"
XX
PN JP2004016117-A.
XX
PD 22-JAN-2004.
XX
PF 17-JUN-2002; 2002JP-00176132.
XX
PR 17-JUN-2002; 2002JP-00176132.
XX
PA (TOYM ) TOYOBO KK.
XX
DR WPI; 2004-113878/12.
DR P-PSDB; ADK68457.
XX
PT New fusion protein useful in diagnosis of diseases and in manufacture of
PT pharmaceutical products, comprises a saccharide binding protein and beta
PT 1, 3-galactosyl transferase.
XX
PS Example 1; SEQ ID NO 1; 23pp; Japanese.
XX
CC This invention relates to a novel recombinant fusion protein that
CC comprises a saccharide binding protein, in particular a maltose binding
CC protein, and a beta-1,3-galactosyl transferase and an appropriate
CC manufacturing method. Specifically, it refers to fusion protein that can
CC transfer galactose to an N-acetyl glucosamine residue or N-acetyl
CC glucosamine monosaccharide. The present invention describes a method to
CC produce this enzyme, cheaply and efficiently, for the diagnosis and
CC treatment of inflammation, infectious diseases or for cancer metastasis
CC suppression. Furthermore, it can also be useful for improving dairy
CC products. Accordingly, compositions exhibit various activities including
CC antiinflammatory, antimicrobial and cytostatic. This polynucleotide
CC sequence is the human beta-1,3-galactosyl transferase cDNA of the
CC invention.
XX
SQ Sequence 933 BP; 227 A; 232 C; 241 G; 233 T; 0 U; 0 Other;

Query Match 30.8%; Score 35.4; DB 12; Length 933;
Best Local Similarity 97.3%; Pred. No. 0.16;
Matches 36; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 79 ATGGCTTCCCGAAGATGAGATTGATGTATATCTGCC 115
DB 1 ATGGCTTCCCGAAGATGAGATTGATGTATATCTGCC 37

RESULT 10
ABQ77404/c
ID ABQ77404 standard; DNA; 78025 BP.
XX
AC ABQ77404;
XX
DT 10-MAY-2003 (first entry)
XX
DE Human SELP DNA.
XX
KW Human; SELP; vascular disease; cardiac; antiarteriosclerotic; stroke;
KW cerebroprotective; gene therapy; coronary artery disease; ischaemia;
KW myocardial infarction; peripheral vascular disease; pulmonary embolism;
KW venous thromboembolism; forensic; paternity testing; GI3790157; gene;
KW SNP; single nucleotide polymorphism; ds.
XX
OS Homo sapiens.
XX

```

FH Key Location/Qualifiers  
FT variation replace(76666.c)  
FT /\*tag= a  
FT /standard\_name= "SNP"  
FT /note= "Single nucleotide polymorphism (ID SELPv8) which  
FT alters the SELP protein from a Thr to Pro at position  
FT 567"  
XX  
XX WO2003016494-A2.  
XX  
XX 27-FEB-2003.  
XX  
XX 16-AUG-2002; 2002WO-US026343.  
XX  
XX 16-AUG-2001; 2001US-0313097P.  
PR 05-OCT-2001; 2001US-0327485P.  
PR 14-DEC-2001; 2001US-00020141.  
XX  
XX (VITI-) VITIVITY INC.  
XX  
XX Mcarthy J, Ableson A;  
XX  
XX WPI; 2003-300617/29.  
DR P-PSDB; ABG74672.  
XX  
XX Identifying a subject as a candidate for a particular course of therapy  
PT to treat a vascular disease or disorder, e.g. stroke, myocardial  
PT infarction or ischemia by determining the identity of the nucleotide  
PT present at specific positions.  
XX  
XX Claim 3; Fig 9; 568pp; English.  
PS  
XX This invention describes a novel method for identifying a subject as a  
CC candidate for a particular course of therapy to treat a vascular disease  
CC or disorder. The method comprises determining the identity of the  
CC nucleotide present at specific positions, or their complements, and  
CC identifying the subject as a candidate for a particular clinical course  
CC of therapy based on the identity of the nucleotide present in that  
CC specific position. The method can be used for identifying a subject who  
CC is a candidate for further diagnostic evaluation of a vascular disease or  
CC disorder and selecting a clinical course of therapy. The products of the  
CC invention have cardiant, antiarteriosclerotic and cerebroprotective  
CC activity and can be used for gene therapy. The methods disclosed are  
CC useful for treating a vascular disease, e.g. atherosclerosis, coronary  
CC artery disease, myocardial infarction, ischaemia, stroke, peripheral  
CC vascular diseases, venous thromboembolism and pulmonary embolism. The DNA  
CC sequences are useful as fingerprint for detecting different individuals  
CC within the same species applicable in forensic studies and paternity  
CC testing. This sequence encodes the human SELP gene represented in  
CC GI3790157 used to illustrate the method of the invention  
XX  
SQ Sequence 78025 BP; 21389 A; 16015 C; 15843 G; 24778 T; 0 U; 0 Other;  
  
Query Match 28.7%; Score 33; DB 8; Length 78025;  
Best Local Similarity 65.8%; Pred. No. 3.7;  
Matches 48; Conservative 0; Mismatches 25; Indels 0; Gaps 0;  
  
QY 15 TAGCATAAACTAGACACATCCTCATGCTTTGAGGCTCTAATCATGATTGTTGCTCTT 74  
Db 5992 TAGCTGTAAACGAGTACTTCTCCACCTTCTCAGTTGTAATTGGGAGACTTTGCTCTT 5933  
  
QY 75 TCAGATGGCTTTC 87  
Db 5932 TCAGATGGCTTTC 5920  
  
RESULT 11  
ABQ69245\_13/C  
Continuation (14 of 31) of ABQ69245 from base 1300001 (Listeria innocua DNA sequence #68  
WP Sequence split into 31 fragments LOCUS ABQ69245 Accession Abq69245  
Fragment Name Begin End  
WP ABQ69245\_00 1 110000  
WP ABQ69245\_01 100001 210000

WP ABQ69245\_02 200001 310000  
WP ABQ69245\_03 300001 410000  
WP ABQ69245\_04 400001 510000  
WP ABQ69245\_05 500001 610000  
WP ABQ69245\_06 600001 710000  
WP ABQ69245\_07 700001 810000  
WP ABQ69245\_08 800001 910000  
WP ABQ69245\_09 900001 1010000  
WP ABQ69245\_10 1000001 1110000  
WP ABQ69245\_11 1100001 1210000  
WP ABQ69245\_12 1200001 1310000  
WP ABQ69245\_13 1300001 1410000  
WP ABQ69245\_14 1400001 1510000  
WP ABQ69245\_15 1500001 1610000  
WP ABQ69245\_16 1600001 1710000  
WP ABQ69245\_17 1700001 1810000  
WP ABQ69245\_18 1800001 1910000  
WP ABQ69245\_19 1900001 2010000  
WP ABQ69245\_20 2000001 2110000  
WP ABQ69245\_21 2100001 2210000  
WP ABQ69245\_22 2200001 2310000  
WP ABQ69245\_23 2300001 2410000  
WP ABQ69245\_24 2400001 2510000  
WP ABQ69245\_25 2500001 2610000  
WP ABQ69245\_26 2600001 2710000  
WP ABQ69245\_27 2700001 2810000  
WP ABQ69245\_28 2800001 2910000  
WP ABQ69245\_29 2900001 3010000  
WP ABQ69245\_30 3000001 3111208  
  
Query Match 27.8%; Score 32; DB 6; Length 110000;  
Best Local Similarity 58.3%; Pred. No. 8.9;  
Matches 56; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
  
QY 16 AGCATAAACTAGACACATCCTCATGCTTTGAGGCTCTAATCATGATTGTTGCTCTT 75  
Db 68725 AGCATAAACTAGACACATGAGAAATGCACCTTGCCTTAATGGTTGGATTATTATGATA 68666  
  
QY 76 CAGATGGCTTTCCTCCGAGATCAGATTGATGATATC 111  
Db 68665 GGCCTGTATGCTGGAATTCGGGATTGGTTTAAATC 68630  
  
RESULT 12  
ABQ67195\_3  
Continuation (4 of 5) of ABQ67195 from base 300001 (Listeria innocua contig DNA sequence  
WP Sequence split into 5 fragments LOCUS ABQ67195 Accession Abq67195  
Fragment Name Begin End  
WP ABQ67195\_0 1 110000  
WP ABQ67195\_1 100001 210000  
WP ABQ67195\_2 200001 310000  
WP ABQ67195\_3 300001 410000  
WP ABQ67195\_4 400001 495269  
  
Query Match 27.8%; Score 32; DB 6; Length 110000;  
Best Local Similarity 58.3%; Pred. No. 8.9;  
Matches 56; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
  
QY 16 AGCATAAACTAGACACATCCTCATGCTTTGAGGCTCTAATCATGATTGTTGCTCTT 75  
Db 90225 AGCATAAACTAGACACATGAGAAATGCACCTTGCCTTAATGGTTGGATTATTATGATA 90284  
  
QY 76 CAGATGGCTTTCCTCCGAGATCAGATTGATGATATC 111  
Db 90285 GGCCTGTATGCTGGAATTCGGGATTGGTTTAAATC 90320  
  
RESULT 13  
ABQ67128  
ID ABQ67128 standard; DNA; 3796 BP.  
XX  
AC ABQ67128;  
XX

DT 28-AUG-2002 (first entry)  
XX Human angiogenesis associated polynucleotide SEQ ID NO 158.  
DE  
XX  
XX Human; angiogenesis; methylation; eye disease; glaucoma; tumour;  
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcers;  
KW macular degeneration; inflammatory bowel disease; Crohn's disease;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiarteriosclerotic; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200246454-A2.  
FN  
XX  
XX 13-JUN-2002.  
PD  
XX  
XX 06-DEC-2001; 2001WO-EP014320.  
PF  
XX  
XX 06-DEC-2000; 2000DE-01061338.  
PR  
XX  
XX (EPIG-) EPIGENOMICS AG.  
PA  
XX  
XX Schacht O;  
PI  
XX  
XX WPI; 2002-500450/53.  
DR  
XX  
XX New nucleic acid fragments from chemically treated angiogenesis-  
PT associated genes, useful for determining methylation status, e.g. in  
PT diagnosis or treatment of cancer.  
PT  
XX  
XX Claim 1; SEQ ID NO 158; 41pp + Sequence Listing; German.  
PS  
XX  
XX The invention relates to a nucleic acid (I) comprising a segment of 18  
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)  
CC having sequences (ABQ66971-ABQ67178) or their complements (I'), also  
CC related oligomers, are used to evaluate the methylation status and/or  
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for  
CC diagnosis and treatment of eye diseases, proliferative retinopathy,  
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,  
CC diabetic retinopathy, macular degeneration caused by neovascularisation,  
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and  
CC Crohn's disease. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 3796 BP; 878 A; 151 C; 1102 G; 1665 T; 0 U; 0 Other;  
  
Query Match 26.8%; Score 30.8; DB 6; Length 3796;  
Best Local Similarity 61.0%; Pred. NO. 8.2;  
Matches 50; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
  
Qy 31 ACATCCTCATGCTTTGAGGTCATATCATGATTTGCTCTTCATGATGCTTTCCG 90  
Db 3311 AAATGTGAAGTATATGTTATTTATCTAGATTTTTTTTTTTAAGATTTTAAATTG 3370  
  
Qy 91 AAGATGAGATGATGATATATCT 112  
Db 3371 GAGGTGGGATGATTTAAATAT 3392  
  
RESULT 14  
ABZ16396  
ID ABZ16396 standard; DNA; 2000 BP.  
XX  
XX ABZ16396;  
AC  
XX  
XX 21-JAN-2003 (first entry)  
DT  
XX  
XX Arabidopsis thaliana stress regulated gene SEQ ID NO 4201.  
DE  
XX  
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.  
XX  
OS Arabidopsis thaliana.

XX WO200216655-A2.  
FN  
XX  
XX 28-FEB-2002.  
PD  
XX  
XX 24-AUG-2001; 2001WO-US026685.  
PF  
XX  
XX 24-AUG-2000; 2000US-0227866P.  
PR  
XX  
XX 26-JAN-2001; 2001US-0264647P.  
PR  
XX  
XX 22-JUN-2001; 2001US-0300111P.  
PR  
XX  
XX (SCRI ) SCRIPPS RES INST.  
PA  
XX  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
PA  
XX  
XX Harper JF, Kreps J, Wang X, Zhu T;  
PI  
XX  
XX WPI; 2002-304127/34.  
DR  
XX  
XX Identifying a stress condition to which a plant cell has been exposed and  
PT producing plants with increased tolerance to these abiotic stresses.  
PT  
XX  
XX Claim 144; SEQ ID NO 4201; 577pp + Sequence Listing; English.  
PS  
XX  
XX The invention relates to identifying a stress condition to which a plant  
CC cell has been exposed, comprising: (a) contacting nucleic acid  
CC representative of expressed polynucleotides in the plant cell with an  
CC array or probes representative of the plant cell genome; and (b)  
CC detecting a profile of expressed polynucleotides in the plant cell  
CC characteristic of a stress response. The method is useful in the  
CC production of transgenic plants, cells and seeds and in producing plants  
CC with increased tolerance to abiotic stress. The present sequence is that  
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used  
CC in methods of the invention. Note: The sequence data for this patent is  
CC not represented in the printed specification but is based on sequence  
CC information supplied to Derwent by the European Patent Office  
XX  
SQ Sequence 2000 BP; 685 A; 313 C; 274 G; 728 T; 0 U; 0 Other;  
  
Query Match 25.6%; Score 30.6; DB 6; Length 2000;  
Best Local Similarity 55.0%; Pred. No. 7.9;  
Matches 60; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
  
Qy 2 CACCTCAGCTCTTAGCATAAACTAGACATCCTCATGCTTTTGGTCTTAATCATG 61  
Db 1077 CACCAATAAATCAAGCAGCACTCGAAGATGATGATGCTTTTCTTCTAATATTG 1136  
  
Qy 62 GATTTTCTTCTTCAGATGCTTTCCGAGATGAGATGATGATATAT 110  
Db 1137 TGGATTGAATTCATTTGGATTGGTTACCAAAAAAAGAGAGATAATATTT 1185  
  
RESULT 15  
ABZ15338/C  
ID ABZ15338 standard; DNA; 2000 BP.  
XX  
XX ABZ15338;  
AC  
XX  
XX 21-JAN-2003 (first entry)  
DT  
XX  
XX Arabidopsis thaliana stress regulated gene SEQ ID NO 3143.  
DE  
XX  
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.  
XX  
OS Arabidopsis thaliana.  
XX  
XX WO200216655-A2.  
FN  
XX  
XX 28-FEB-2002.  
PD  
XX  
XX 24-AUG-2001; 2001WO-US026685.  
PF  
XX  
XX 24-AUG-2000; 2000US-0227866P.  
PR  
XX  
XX 26-JAN-2001; 2001US-0264647P.  
PR

```

PR 22-JUN-2001; 2001US-0300111P.
XX
XX (SCEI ) SCRIPPS RES INST.
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Harper JF, Krebs J, Wang X, Zhu T;
XX
XX WPI; 2002-304127/34.
DR
XX
XX Identifying a stress condition to which a plant cell has been exposed and
PT producing plants with increased tolerance to these abiotic stresses.
XX
XX Claim 144; SEQ ID NO 3143; 577pp + Sequence Listing; English.
XX
XX The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising: (a) contacting nucleic acid
CC representative of expressed polynucleotides in the plant cell with an
CC array or probes representative of the plant cell genome; and (b)
CC detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used
CC in methods of the invention. Note: The sequence data for this patent is
CC not represented in the printed specification but is based on sequence
CC information supplied to Derwent by the European Patent Office
XX
XX SQ Sequence 2000 BP; 563 A; 407 C; 385 G; 645 T; 0 U; 0 Other;
Query Match 26.4%; Score 30.4; DB 6; Length 2000;
Best Local Similarity 59.1%; Pred. No. 9.2;
Matches 52; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY 19 ATAAACTAGACATCCTCATGCTTTTGAGGCTTAATCATTCATTGTTTGTCTTCAG 78
Db 203 ATTTATCCTGTCTAATCTCATTTTACCTGTTTGATCCTTGGGTTTGGCTTCTCTAT 144
QY 79 ATGGCTTCCCGAGATGAGATGATGT 106
Db 143 GTCGGCTTTTGGGAGATTGGTTGATTT 116

```

Search completed: September 22, 2005, 12:38:51  
Job time : 72.5965 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2005, 12:09:12 ; Search time 507.076 Seconds  
(without alignments)  
8632.613 Million cell updates/sec

Title: US-10-777-828-8\_COPY\_1\_115

Perfect score: 115

Sequence: 1 ccacctcagctctagcat.....gagattgatctatctgcc 115

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gsal:\*  
9: gb\_gsal2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35.4	30.8	933	AY419460	Homo sapi
2	34	29.6	263	CB444598	tigr-ges
3	33.8	29.4	780	AY419461	Pan trogl
4	33.4	29.0	848	AQ750178	HS_5575_A
5	32.8	28.5	436	CA834817	MCS031A04
6	32.6	28.3	879	BU395580	603804144
7	32.2	28.0	735	CR248164	Reverse s
8	31.8	27.7	701	BH932547	odh46f07.
9	31.8	27.7	839	CG457920	PURXW21TB
10	31.8	27.7	889	CG169346	PURFE58TD
11	31.6	27.5	370	BE110064	UI-R-CA0-
12	31.4	27.3	685	BB275595	BB275595
13	31.2	27.1	424	CK515682	rsbjb0_00
14	31.2	27.1	541	B0109768	imageqc_7
15	31.2	27.1	584	RJ087028	EJ087028
16	31	27.0	492	CP272903	EST2465_Z
17	30.8	26.8	630	CB256589	70-E01079
18	30.8	26.8	760	BF794004	602254647
19	30.6	26.6	386	AU227311	AU227311
20	30.6	26.6	425	AU236398	AU236398
21	30.6	26.6	525	AQ370569	HS_5043_A
22	30.6	26.6	598	BZ484243	BOOAB11TR
23	30.6	26.6	614	AZ319319	1M0038C20
24	30.6	26.6	927	BX451934	BX451934

25	30.4	26.4	639	4	BJ488633	BJ488633
c 26	30.4	26.4	729	4	BJ748272	BJ748272
c 27	30.4	26.4	736	4	BJ509194	BJ509194
28	30.4	26.4	1174	6	CB561585	AGENCOURT
29	30.2	26.3	226	1	AV111150	AV111150
30	30.2	26.3	240	1	AJ484636	AJ484636
31	30.2	26.3	331	7	CO325225	EK195856
32	30.2	26.3	360	1	AJ484637	AJ484637
c 33	30.2	26.3	779	8	BZ270094	CH230-275
34	30.2	26.3	1018	8	CC204655	CH261-152
35	30.2	26.3	1136	8	CC184502	CH261-820
36	30	26.1	541	9	CR076191	Forward s
c 37	30	26.1	633	9	CG107952	PUIJAS8TD
c 38	30	26.1	668	7	CV472267	45579.1 C
c 39	30	26.1	742	9	AG414481	Mus muscu
c 40	30	26.1	792	8	BH951588	odh16b07.
41	30	26.1	820	9	CR024687	Forward s
c 42	29.8	25.9	224	1	AA810738	082801.8
c 43	29.8	25.9	235	1	A1004187	OU56D11.X
c 44	29.8	25.9	266	1	A1524537	t164B05.X
c 45	29.8	25.9	347	1	AA824559	OC78G05.8

#### ALIGNMENTS

RESULT 1  
AY419460  
LOCUS  
DEFINITION Homo sapiens B3GALT5 gene, VIRTUAL TRANSCRIPT, partial sequence,  
AY419460  
ACCESSION  
VERSION AY419460.1 GI:39775417  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 933)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J., Adams,M.D. and Cargill,M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 933)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J., Adams,M.D. and Cargill,M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA  
COMMENT These sequences were made by sequencing genomic exons and ordering them based on alignment.  
FEATURES  
Location/Qualifiers  
1..933  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
<1..>933  
/gene="B3GALT5"  
/locus\_tag="HCM6898"

Query Match 30.8%; Score 35.4; DB 9; Length 933;  
Best Local Similarity 97.3%; Pred. No. 3.5;  
Matches 36; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 79 ATGCTTTCCGAGATGAGATTGATGATATCTGCC 115  
|||||

```

Db      1  ATGGCTTTCCGAAGATGAGATTCGATGATATTTGCC 37

RESULT 2
CE444598
LOCUS    tigr-gss-dog-17000336367589 Dog Library Canis familiaris genomic,
DEFINITION 263 bp DNA linear GSS 27-SEP-2003
genomic survey sequence.
ACCESSION CE444598
VERSION    CE444598.1 GI:36729674
KEYWORDS   GSS.
SOURCE     Canis familiaris (dog)
ORGANISM   Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1  (bases 1 to 263)
AUTHORS    Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
            The dog genome: survey sequencing and comparative analysis
            Science 301 (5641), 1898-1903 (2003)
            22875432
            14512627
            Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source           1..263
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site 1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN
Query Match      29.6%; Score 34; DB 9; Length 263;
Best Local Similarity 63.4%; Pred. No. 7.8;
Matches 52; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy      27  AGACACATCCTTCATGCTTTTGAGGTCTAATCATTCGATTTTGTCCTTCAGATGGCTTT 86
        |||||
        1  AAACGGTCATCATGCTCTTCGCTCTTTGTCCTTCATGTTGTCCTTTTAGTCTGTTT 60
        |||||

Qy      87  CCCGAAGATGAGATTCGATGAT 108
        |||||
Db      61  GTTTAAGATTTTATTATGATAT 82

RESULT 3
AY419461
LOCUS    Pan troglodytes B3GALT5 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY419461
VERSION    AY419461.1 GI:39775418
KEYWORDS   GSS.
SOURCE     Pan troglodytes (chimpanzee)
ORGANISM   Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1  (bases 1 to 780)
AUTHORS    Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
            Inferring nonneural evolution from human-chimp-mouse orthologous
            .
            .
TITLE

```

```

gene trios
Science 302 (5652), 1960-1963 (2003)
14671302
2  (bases 1 to 780)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES             Location/Qualifiers
     source           1..780
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     <1..>780
                     /gene="B3GALT5"
                     /locus_tag="HCM6898"
ORIGIN
Query Match      29.4%; Score 33.8; DB 9; Length 780;
Best Local Similarity 94.6%; Pred. No. 11;
Matches 35; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      79  ATGGCTTTCCGAAGATGAGATTCGATGATATTCGTC 115
        |||||
        1  ATGGCTTTCCGAAGATGAGATTCGATGATATTCGTC 37
        |||||

RESULT 4
AQ750178
LOCUS    HS 5575_A2 G10 SP6 RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic_clone Plate=1151 Col=20 Row=M, genomic survey sequence.
ACCESSION AQ750178
VERSION    AQ750178.1 GI:5537336
KEYWORDS   GSS.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1  (bases 1 to 848)
AUTHORS    Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
            Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
            Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
            99380589
            10449764
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htcsc.washington.edu
            Plate: 1151 row: M column: 20
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 848.
            Location/Qualifiers
            1..848
            /organism="Homo sapiens"
FEATURES             Location/Qualifiers
     source           1..848

```

ORIGIN

Query Match 28.3%; Score 32.6; DB 5; Length 879;  
Best Local Similarity 58.9%; Pred. NO. 26;  
Matches 56; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```
Qy 11 CTCCTAGCATAAACTAGACACATCTCATGCTTTTGAGGTCTAATCATTTGGATTGTT 70
Db 292 CTCGTATAAGACAACCTGCACCTAGCATTAATGGTCTATAGATCATGCTCCTTAGCTTAGTA 351
Qy 71 CTTTTCAGATGGCTTTCCCGAAGATGAGATTGATG 105
Db 352 CCTAATACTTAGCAGTGTCATCAGGTGAGATTGATG 386

RESULT 7
CR248164 735 bp DNA linear GSS 06-JUL-2004
LOCUS Reverse strand read from insert in 5'HPRT insertion targeting and
DEFINITION chromosome engineering clone MHPN85b22, genomic survey sequence.
ACCESSION CR248164
VERSION GI:50027018
KEYWORDS GSS; genome survey sequence; MICBR.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 735)
AUTHORS Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
Jongkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
Rogers,J. and Bradley,A.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. http://www.sanger.ac.uk/MICR
FEATURES
source
location/Qualifiers
1..735
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="MHPN85b22"
/clone_lib="MHPN"

ORIGIN
Query Match 28.0%; Score 32.2; DB 9; Length 735;
Best Local Similarity 57.4%; Pred. No. 34;
Matches 58; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 15 TAGCATAAACTAGACACATCTCATGCTTTTGAGGTCTAATCATTTGGATTGTTCTCTT 74
Db 443 TAGAAAATGCTACTCTCTATATATTAAGTGACTTTTCTCAATGGGTTTCTCTCTT 502
Qy 75 TCAGATGGCTTTCCCGAAGATGAGATTGATGTATATCTGCC 115
Db 503 TCAGATGACTCTAGCTAATATATCATGTTGCACATGAACTGTC 543

RESULT 8
BH932547 701 bp DNA linear GSS 01-OCT-2002
LOCUS odb46f07.g1 B.oleracea002 Brassica oleracea genomic, genomic survey
DEFINITION sequence.
ACCESSION BH932547
VERSION BH932547.1 GI:23412613
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 701)
AUTHORS Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,
Nash,W., Rabinowicz,P.D. and Wilson,R.K.
TITLE Whole genome shotgun reads from Brassica oleracea
JOURNAL Unpublished (2002)
CONTACT: Richard K. Wilson
COMMENT Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: odb46 row: f column: 07
```

```
Seq primer: -28RppOT reverse
Class: shotgun
High quality sequence start: 16
High quality sequence stop: 217.
FEATURES
source
location/Qualifiers
1..701
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="vector: pOTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea T01000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."

ORIGIN
Query Match 27.7%; Score 31.8; DB 8; Length 701;
Best Local Similarity 64.0%; Pred. No. 44;
Matches 48; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 14 CTAGCATAAACTAGACACATCTCTCATGCTTTTGAGGTCTAATCATTTGGATTGTTCTCT 73
Db 315 CGATTATGAACTTTTAACATCTTCATCTGTTGGGTATTGTTATTGCTCTTTGGTGCT 374
Qy 74 TTCAGATGGCTTTCC 88
Db 375 TGATGAAGTGTTTCC 389

RESULT 9
CG457920 839 bp DNA linear GSS 17-SEP-2003
LOCUS PUFXM23TBC ZM 0.6 1.0 KB Zea mays genomic clone ZMMBta0755D21,
DEFINITION genomic survey sequence.
ACCESSION CG457920
VERSION CG457920.1 GI:34842920
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 839)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other GSSs: PUFXM23TDC
Contact: Cathy Whitelaw
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES
source
location/Qualifiers
1..839
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBta0755D21"
/clone_lib="ZM 0.6 1.0 KB"
/note="vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
Cor selected genomic DNA library"

ORIGIN
Query Match 27.7%; Score 31.8; DB 9; Length 839;
```

```

Best Local Similarity 56.1%; Pred. No. 46;
Matches 60; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 6 TCAGCCTCCTAGCATAAACATAGACACATCCTCATGCTTTTGAGGTCTAATCATTTGGATT 65
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31 TGAAGTTATATGTAATCCATTCATTAACCTTAGGTTAGAAACACATTAATTCATT 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 66 TTGTTCCCTTCAGATGGCTTTCCCGAAGATGAGATTGATGATATCT 112
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 TTATTCCCTTTAGGATAAACTAAGTCATATATATATATATATATATAT 137
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
CGI69346 889 bp DNA linear GSS 21-AUG-2003
PUJFE56TD ZM 0.6-1.0 KB Zea mays genomic clone ZMWBTA0654J15,
Genomic survey sequence.
ACCESSION CGI69346
KEYWORDS CGI69346.1 GI:34060147
SOURCE GSS.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 889)
White, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Renick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Benner, J.
Maize Genomics Consortium
Unpublished (2003)
Other GSSs: PUJFE56TB
Contact: Cathy Whitelaw
TIGR
712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.

FEATURES
    source
    1..889
        /organism="Zea mays"
        /mol_type="genomic DNA"
        /strain="B73"
        /db_xref="taxon:4577"
        /clone="ZMWBTA0654J15"
        /note="Vector: PCR4-TOPO; site_1: EcoRI; 0.6-1.0 kb high
        Cot selected genomic DNA library"

ORIGIN
Query Match 27.7%; Score 31.8; DB 9; Length 889;
Best Local Similarity 56.1%; Pred. No. 46;
Matches 60; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 6 TCAGCCTCCTAGCATAAACATAGACACATCCTCATGCTTTTGAGGTCTAATCATTTGGATT 65
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 320 TGAAGTTATATGTAATCCATTCATTAACCTTAGGTTAGAAACACATTAATTCATT 379
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 66 TTGTTCCCTTCAGATGGCTTTCCCGAAGATGAGATTGATGATATCT 112
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 380 TTATTCCCTTTAGGATAAACTAAGTCATATATATATATATATATATAT 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
BE110064/c 370 bp mRNA linear EST 13-JUN-2000
LOCUS BE110064
DEFINITION UI-R-CA0-axj-f-03-0-UI-s1 UI-R-CA0 Rattus norvegicus cDNA clone
ACCESSION BE110064
VERSION BE110064.1 GI:8502169
KEYWORDS EST.

```

```

SOURCE
ORGANISM Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 370)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRP, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized hippocampus library cDNA Library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA-Yes.
    Location/Qualifiers
    1..370
        /organism="Rattus norvegicus"
        /mol_type="mRNA"
        /strain="Sprague-Dawley"
        /db_xref="taxon:10116"
        /clone="UI-R-CA0-axj-f-03-0-UI"
        /lab_host="DH10B (Life Technologies)"
        /clone_lib="UI-R-CA0"
        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
        polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CA0
        library is a subtracted library derived from the following
        tissues: thalamus, cerebellum, hypothalamus, medulla,
        pons, midbrain, cerebral cortex, corpus striatum, testis,
        and hippocampus. For a detailed description of the
        library from which this clone was derived, please visit
        our web site at ratest.eng.uiowa.edu. The subtraction
        has been previously described in (Bonaldo, Lennon and
        Soares, Genome Research 6:791-806, 1996)
        TAG LIB=UI-R-CA0
        TAG_SEQ=GATTG"

ORIGIN
Query Match 27.5%; Score 31.6; DB 2; Length 370;
Best Local Similarity 60.5%; Pred. No. 46;
Matches 52; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 7 CAGCCTCCTAGCATAAACATAGACACATCCTCATGCTTTTGAGGTCTAATCATTTGGATT 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128 CTGCCACTCAGCAAGACTGGGTACATTCATGATGTTGATTTCTTCTTGTGGCTTC 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 67 TGTTCCTTTTCAGATGGCTTTCCCGAA 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 ATTTCCTTACTCAGTGGCTCCCGAA 43
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
BB275595 685 bp mRNA linear EST 24-OCT-2001
LOCUS BB275595
DEFINITION BB275595 RIKEN full-length enriched, 10 days neonate cortex Mus
ACCESSION BB275595
VERSION BB275595.2 GI:16400850

```



```
QY 15 TAGCATAAACTAGACACATCCTCATGCTTTTGAGGTCATATCATGATTTGTTCCCTT 74
Db 375 TAGCGTACATCTCGAATCTTACAGGTTCTGGAATCCAAACATTTGTTTGAATCGTT 316

QY 75 TCAGATGCTGCTTCCCGAAGATGAGATTGATG 106
Db 315 TCGCCCACTGACCAAGATTGAGGTGATTT 284

RESULT 14
LOCUS BQ109768 541 bp mRNA linear EST 16-APR-2002
DEFINITION imageqc_7_2001/snn459bdf41.xl NICHD_XGC_Embi Xenopus laevis cDNA
clone IMAGE:5161929 3', mRNA sequence.
ACCESSION BQ109768
VERSION BQ109768.1 GI:20159422
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
REFERENCE 1 (bases 1 to 541)
AUTHORS Kale, P.I., Harsch, T.J., Foltz, P.A., Nelson, D.O., Sanders, C.G. and
Prange, C.K.
TITLE The I.M.A.G.E. Consortium quality control effort: clone
resequencing for verification
JOURNAL Unpublished (2001)
COMMENT Contact: Prange CK
The I.M.A.G.E. Consortium
Lawrence Livermore National Laboratory
Livermore, CA, USA
Email: help@image.llnl.gov
This read has been produced as part of the I.M.A.G.E. Consortium
quality control effort. High quality sequence is defined as having
100 or more base pairs with a phred quality value of 20 or greater,
where a sliding window of 4 base pairs with a phred quality value
of 15 or greater marks the beginning and end of the sequence. For
information on obtaining this clone, please contact
info@image.llnl.gov.
Plate: LLAM11401 row: n column: 10
Seq primer: -21m3
High quality sequence stop: 541.
Location/Qualifiers
source
1..541
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:5161929"
/tissue type="embryo (stage 10)"
/lab host="DH10B (phage-resistant)"
/clone_lib="NICHD_XGC_Emb1"
Note="Vector: pCMV-SF0RT6; Site 1: NotI; Site 2: SalI;
Cloned unidirectionally. Primer: Oligo dt. Average insert
size 1.55 kb. Constructed by Life Technologies. Note: This
is a Xenopus Gene Collection (XGC) library."

ORIGIN
Query Match 27.1%; Score 31.2; DB 5; Length 541;
Best Local Similarity 58.7%; Pred. No. 65;
Matches 54; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 18 CATAAACTAGACACATCCTCATGCTTTTGAGGTCATATCATGATTTGTTCCCTTCA 77
Db 442 CATAGATTATTACAATACTAGTGCTTTCAGGATTGAATTAGATTTTGTGCTT 383

QY 78 GATGCTTTCCCGAAGATGAGATTGATGTATA 109
Db 382 AATAGCATTTACGGAAATAGATTATGTATA 351

RESULT 15
LOCUS BQ087028 584 bp mRNA linear EST 29-SEP-2003
DEFINITION BQ087028 NIBB Mochii normalized Xenopus tailbud library Xenopus
laevis cDNA clone XL055d11 3', mRNA sequence.
ACCESSION BQ087028
VERSION BQ087028.1 GI:17583656
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
REFERENCE 1 (bases 1 to 584)
AUTHORS Kikayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and
Kohara, Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tehinigenes.nig.ac.jp
The information of this clone is available through the following
URL.
http://xenopus.nibb.ac.jp.
Location/Qualifiers
source
1..584
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL055d11"
/tissue type="whole embryo"
/dev stage="stage 25"
/clone_lib="NIBB Mochii normalized Xenopus tailbud
library"

ORIGIN
Query Match 27.1%; Score 31.2; DB 4; Length 584;
Best Local Similarity 58.7%; Pred. No. 66;
Matches 54; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 18 CATAAACTAGACACATCCTCATGCTTTTGAGGTCATATCATGATTTGTTCCCTTCA 77
Db 430 CATAGATTATTACAATACTAGTGCTTTCAGGATTGAATTAGATTTTGTGCTT 371

QY 78 GATGCTTTCCCGAAGATGAGATTGATGTATA 109
Db 370 AATAGCATTTACGGAAATAGATTATGTATA 339

Search completed: September 22, 2005, 16:56:02
Job time : 516.076 secs
```

his Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2005, 12:15:17 ; Search time 23.9415 Seconds  
(without alignments)  
7859.644 Million cell updates/sec

Title: US-10-777-828-8\_COPY\_1\_115

Perfect score: 115

Sequence: 1 ccacctcagctcctagcat.....gagattgatgtatctgcc 115

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115	100.0	1011	4	US-09-831-630-8
2	30.2	26.3	327	4	US-09-270-767-2980
3	30.2	26.3	327	4	US-09-270-767-18262
4	30	26.1	99960	4	US-09-762-311-2
5	29	25.2	359	4	US-08-958-171E-2970
6	29	25.2	359	4	US-08-781-986A-2970
7	28.8	25.0	2330	3	US-09-120-653D-4
8	28.8	25.0	3910	3	US-09-120-653D-1
9	28.4	24.7	524032	4	US-09-949-016-16928
10	28.4	24.7	524032	4	US-09-949-016-16929
11	28.4	24.7	524032	4	US-09-949-016-16930
12	28.4	24.7	524032	4	US-09-949-016-16931
13	28.4	24.7	529885	4	US-09-949-016-14340
14	28.4	24.7	529885	4	US-09-949-016-14341
15	28.4	24.7	529885	4	US-09-949-016-14342
16	28.4	24.7	529885	4	US-09-949-016-14343
17	28.4	24.7	529885	4	US-09-949-016-14344
18	28.4	24.7	529885	4	US-09-949-016-14345
19	28.4	24.7	529885	4	US-09-949-016-14346
20	28.4	24.7	529885	4	US-09-949-016-14347
21	28.2	24.5	66490	4	US-09-949-016-15849
22	28.2	24.2	10321	4	US-09-949-016-13587
23	27.6	24.0	601	4	US-09-949-016-149681
24	27.6	24.0	640681	3	US-09-790-988-1
25	27.4	23.8	513	3	US-09-188-930-237
26	27.4	23.8	513	4	US-09-312-283C-237
27	27.4	23.8	2124	4	US-09-328-352-708

28	27.4	23.8	40936	4	US-09-949-016-16607	Sequence 16607, A
29	27.4	23.8	40936	4	US-09-949-016-16608	Sequence 16608, A
c 30	27.4	23.8	205163	4	US-09-949-016-17009	Sequence 17009, A
c 31	27.2	23.7	601	4	US-09-949-016-51191	Sequence 51191, A
32	27.2	23.7	17226	4	US-09-949-016-13806	Sequence 13806, A
c 33	27.2	23.7	118923	4	US-09-949-016-13227	Sequence 13227, A
c 34	27.2	23.7	153866	4	US-09-949-016-16919	Sequence 16919, A
c 35	27.2	23.7	235064	4	US-09-949-016-15390	Sequence 15390, A
36	27.2	23.7	276687	4	US-09-949-016-13840	Sequence 13840, A
37	27	23.5	21234	3	US-09-810-671-3	Sequence 3, Appli
38	27	23.5	21234	4	US-10-109-854-3	Sequence 3, Appli
39	27	23.5	21234	4	US-10-339-656-3	Sequence 3, Appli
c 40	27	23.5	92505	4	US-09-949-016-14018	Sequence 14018, A
41	27	23.5	130971	4	US-09-949-016-14205	Sequence 14205, A
42	27	23.5	156324	4	US-09-949-016-13749	Sequence 13749, A
c 43	27	23.5	390416	4	US-09-949-016-16923	Sequence 16923, A
c 44	26.8	23.3	601	4	US-09-949-016-44719	Sequence 44719, A
c 45	26.8	23.3	22390	4	US-09-949-016-15311	Sequence 15311, A

#### ALIGNMENTS

##### RESULT 1

US-09-831-630-8  
; Sequence 8, Application US/09831630  
; Patent No. 6800468  
; GENERAL INFORMATION:  
; APPLICANT: Clausen, Henrik  
; APPLICANT: Amado, Margarita  
; TITLE OF INVENTION: UDP-GALACTOSE: BETA-N-ACETYL-GLUCOSAMINE BETA 1,3  
; FILE REFERENCE: 7188-157  
; CURRENT APPLICATION NUMBER: US/09/831,630  
; CURRENT FILING DATE: 2001-07-12  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 1011  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (79)..(1008)  
US-09-831-630-8

Query Match	100.0%	Score 115;	DB 4;	Length 1011;
Best Local Similarity	100.0%	Pred. No. 3.2e-31;		
Matches 115;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	CCACCTCAGCTCCTAGCATAAACTAGACACATCCCTCATCGCTTTTGAGGCTTAATCATTT	60	
Db	1	CCACCTCAGCTCCTAGCATAAACTAGACACATCCCTCATCGCTTTTGAGGCTTAATCATTT	60	
Qy	61	GGATTTTGTTCCTTTCAGATGGCTTTCCGAGATGAGATTGATGATATCTGCC	115	
Db	61	GGATTTTGTTCCTTTCAGATGGCTTTCCGAGATGAGATTGATGATATCTGCC	115	
RESULT 2				
US-09-270-767-2980				
; Sequence 2980, Application US/09270767				
; Patent No. 6703491				
; GENERAL INFORMATION:				
; APPLICANT: Homburger et al.				
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster				
; FILE REFERENCE: File Reference: 7326-094				
; CURRENT APPLICATION NUMBER: US/09/270,767				
; CURRENT FILING DATE: 1999-03-17				
; NUMBER OF SEQ ID NOS: 62517				
; SOFTWARE: Patent In Ver. 2.0				
; SEQ ID NO 2980				
; LENGTH: 327				

```
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-2980

Query Match      26.3%; Score 30.2; DB 4; Length 327;
Best Local Similarity 56.6%; Pred. No. 0.49;
Matches 56; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 2 CACCTCAGCTCCCTAGCATAAACATAGACACATCCTCATGCTTTTGAGGTCTAATCATTTG 61
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 163 CACCTAATACTAGTAGCAATAATGTAAGACAGACAAATGCAGATAATGAATTCCTCATTT 222
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 62 GATTTTGTCTTTTCAGATGGCTTTCCCGAAGATGAGAT 100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 223 CATTTAATTCAAAGCAAAACCGATTTCCAAGAGGGCAAT 261
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
US-09-270-767-18262
; Sequence 18262, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18262
; LENGTH: 327
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-18262

Query Match      26.3%; Score 30.2; DB 4; Length 327;
Best Local Similarity 56.6%; Pred. No. 0.49;
Matches 56; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 2 CACCTCAGCTCCCTAGCATAAACATAGACACATCCTCATGCTTTTGAGGTCTAATCATTTG 61
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 163 CACCTAATACTAGTAGCAATAATGTAAGACAGACAAATGCAGATAATGAATTCCTCATTT 222
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 62 GATTTTGTCTTTTCAGATGGCTTTCCCGAAGATGAGAT 100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 223 CATTTAATTCAAAGCAAAACCGATTTCCAAGAGGGCAAT 261
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
US-09-762-311-2/C
; Sequence 2, Application US/09762311
; Patent No. 6825004
; GENERAL INFORMATION:
; APPLICANT: BLUMENFELD, Marta
; APPLICANT: BOUGUELERET, Lydie
; APPLICANT: CHUMAKOV, Ilya
; TITLE OF INVENTION: Nucleic Acids Encoding Human TBC-1 Protein And Polymorphic Markers
; FILE REFERENCE: 46.US2.PCT
; CURRENT APPLICATION NUMBER: US/09/762,311
; CURRENT FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: 2001-06-21
; PRIOR FILING DATE: 1998-08-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent.pm
; SEQ ID NO 2
; LENGTH: 99960
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 4661..4789
; OTHER INFORMATION: exon A
; NAME/KEY: exon
; LOCATION: 6116..6202
; OTHER INFORMATION: exon B
; NAME/KEY: exon
; LOCATION: 9919..10199
; OTHER INFORMATION: exon C
; NAME/KEY: exon
; LOCATION: 14521..14660
; OTHER INFORMATION: exon D
; NAME/KEY: exon
; LOCATION: 50257..50442
; OTHER INFORMATION: exon E
; NAME/KEY: exon
; LOCATION: 56256..56417
; OTHER INFORMATION: exon F
; NAME/KEY: exon
; LOCATION: 63326..63484
; OTHER INFORMATION: exon G
; NAME/KEY: exon
; LOCATION: 76036..76280
; OTHER INFORMATION: exon H
; NAME/KEY: exon
; LOCATION: 78364..78523
; OTHER INFORMATION: exon I
; NAME/KEY: exon
; LOCATION: 85295..85464
; OTHER INFORMATION: exon J
; NAME/KEY: exon
; LOCATION: 93417..93590
; OTHER INFORMATION: exon K
; NAME/KEY: exon
; LOCATION: 97476..97960
; OTHER INFORMATION: exon L
; NAME/KEY: misc.feature
; LOCATION: 97961..99960
; OTHER INFORMATION: 3' regulatory region
; NAME/KEY: allele
; LOCATION: 1443
; OTHER INFORMATION: 99-20508-456 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 5247
; OTHER INFORMATION: 99-20469-213 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 6223
; OTHER INFORMATION: 5-254-227 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 14723
; OTHER INFORMATION: 5-257-353 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 19186
; OTHER INFORMATION: 99-20511-32 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 18997
; OTHER INFORMATION: 99-20511-221 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 19891
; OTHER INFORMATION: 99-20510-115 : deletion of TCT
; NAME/KEY: allele
; LOCATION: 29617
; OTHER INFORMATION: 99-20504-90 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 42519
; OTHER INFORMATION: 99-20493-238 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 69324
; OTHER INFORMATION: 99-20499-221 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 69181
; OTHER INFORMATION: 99-20499-364 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 69146
; OTHER INFORMATION: 99-20499-399 : polymorphic base A or G
; NAME/KEY: allele
```

LOCATION: 76458  
OTHER INFORMATION: 99-20473-138 : deletion of TAACA  
NAME/KEY: allele  
LOCATION: 78595  
OTHER INFORMATION: 5-249-304 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 82159  
OTHER INFORMATION: 99-20485-269 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 84522  
OTHER INFORMATION: 99-20481-131 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 84810  
OTHER INFORMATION: 99-20481-419 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 89967  
OTHER INFORMATION: 99-20480-233 : polymorphic base A or G  
NAME/KEY: primer\_bind  
LOCATION: 988..1006  
OTHER INFORMATION: 99-20508.pu  
NAME/KEY: primer\_bind  
LOCATION: 1509..1529  
OTHER INFORMATION: 99-20508.rp complement  
NAME/KEY: primer\_bind  
LOCATION: 5039..5056  
OTHER INFORMATION: 99-20469.pu  
NAME/KEY: primer\_bind  
LOCATION: 5534..5554  
OTHER INFORMATION: 99-20469.rp complement  
NAME/KEY: primer\_bind  
LOCATION: 5997..6015  
OTHER INFORMATION: 5-254.pu  
NAME/KEY: primer\_bind  
LOCATION: 6332..6350  
OTHER INFORMATION: 5-254.rp complement  
NAME/KEY: primer\_bind  
LOCATION: 14371..14390  
OTHER INFORMATION: 5-257.pu  
NAME/KEY: primer\_bind  
LOCATION: 14798..14817  
OTHER INFORMATION: 5-257.rp complement  
NAME/KEY: primer\_bind  
LOCATION: 18751..18771  
OTHER INFORMATION: 99-20511.rp  
NAME/KEY: primer\_bind  
LOCATION: 19198..19217  
OTHER INFORMATION: 99-20511.pu complement  
NAME/KEY: primer\_bind  
LOCATION: 19605..19625  
OTHER INFORMATION: 99-20510.rp  
NAME/KEY: primer\_bind  
LOCATION: 19986..20005  
OTHER INFORMATION: 99-20510.pu complement  
NAME/KEY: primer\_bind  
LOCATION: 29529..29547  
OTHER INFORMATION: 99-20504.pu  
NAME/KEY: primer\_bind  
LOCATION: 30041..30061  
OTHER INFORMATION: 99-20504.rp complement  
NAME/KEY: primer\_bind  
LOCATION: 42268..42287  
OTHER INFORMATION: 99-20493.rp  
NAME/KEY: primer\_bind  
LOCATION: 42732..42752  
OTHER INFORMATION: 99-20493.pu complement  
NAME/KEY: primer\_bind  
LOCATION: 69026..69046  
OTHER INFORMATION: 99-20499.rp  
NAME/KEY: primer\_bind  
LOCATION: 69525..69543  
OTHER INFORMATION: 99-20499.pu complement  
NAME/KEY: primer\_bind  
LOCATION: 76323..76343

OTHER INFORMATION: 99-20473.pu  
NAME/KEY: primer\_bind  
LOCATION: 76771..76790  
OTHER INFORMATION: 99-20473.rp complement  
NAME/KEY: primer\_bind  
LOCATION: 78292..78309  
OTHER INFORMATION: 5-249.pu  
NAME/KEY: primer\_bind  
LOCATION: 78704..78721  
OTHER INFORMATION: 5-249.rp complement  
NAME/KEY: primer\_bind  
LOCATION: 81893..81912  
OTHER INFORMATION: 99-20485.pu  
NAME/KEY: primer\_bind  
LOCATION: 82353..82372  
OTHER INFORMATION: 99-20485.rp complement  
NAME/KEY: primer\_bind  
LOCATION: 84392..84412  
OTHER INFORMATION: 99-20481.pu  
NAME/KEY: primer\_bind  
LOCATION: 84909..84929  
OTHER INFORMATION: 99-20481.rp complement  
NAME/KEY: primer\_bind  
LOCATION: 89746..89765  
OTHER INFORMATION: 99-20480.rp  
NAME/KEY: primer\_bind  
LOCATION: 90179..90198  
OTHER INFORMATION: 99-20480.pu complement  
NAME/KEY: primer\_bind  
LOCATION: 9475..9493  
OTHER INFORMATION: 99-430-352.mis  
Query Match 26.1%; Score 30; DB 4; Length 99960;  
Best Local Similarity 54.5%; Pred No. 6.4;  
Matches 60; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
QY 1 CCACCTCAGCCTCCTAGCATAAACAGACACATCCTCATGCTTTTGAGGTCAATCAT 60  
Db 90492 CCAGCTCAACCTACCATCATATTAATCCAGTCATCAAGTGTAGACATCTAGTCCA 90433  
QY 61 GGATTTGTTCTTCAGATGGCTTCCGGAAGATGAGATGATGATATAT 110  
Db 90432 AGTTCTGTTATTATAGTGACAAAGTCTAGGAGATATCCTTATACATAT 90383  
RESULT 5  
US-08-956-171E-2970  
; Sequence 2970, Application US/08956171E  
; Patent No. 6593114  
GENERAL INFORMATION:  
APPLICANT: Charles Kunesch  
Gil H. Choi  
Patrick S. Dillon  
Craig A. Rosen  
Steven C. Barash  
Michael R. Fannon  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5256  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/956,171E  
FILING DATE: 20-Oct-1997

```
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 2970:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2970:
US-08-956-171E-2970

Query Match      25.2%; Score 29; DB 4; Length 359;
Best Local Similarity 54.9%; Pred. No. 1.4;
Matches 56; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 10 CCTCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGTCTAATCATTTGGATTGTTGT 69
DB 231 CATCATACCATATAACTTTTATCATATAATCATTTTCAGCGAACTTTAGTTTGATGTTTTC 290

QY 70 TCCTTTTCAGATGGCTTCCCGAAGATGAGATTGATGTATATC 111
DB 291 TTGATTAAATCTTTTCGGCGANCCCTCAGCTTGATGTTTTC 332

RESULT 6
US-08-781-986A-2970
; Sequence 2970, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 2970:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 base pairs
; TYPE: nucleic acid
```

```
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-2970

Query Match      25.2%; Score 29; DB 4; Length 359;
Best Local Similarity 54.9%; Pred. No. 1.4;
Matches 56; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 10 CCTCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGTCTAATCATTTGGATTGTTGT 69
DB 231 CATCATACCATATAACTTTTATCATATAATCATTTTCAGCGAACTTTAGTTTGATGTTTTC 290

QY 70 TCCTTTTCAGATGGCTTCCCGAAGATGAGATTGATGTATATC 111
DB 291 TTGATTAAATCTTTTCGGCGANCCCTCAGCTTGATGTTTTC 332

RESULT 7
US-09-120-653D-4
; Sequence 4, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
; APPLICANT: YOON, JI-WON
; APPLICANT: JUN, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: AHN, JONG-SEONG
; APPLICANT: HA, YOUNG-JU
; APPLICANT: CHUNG, SOO-IL
; TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
; FILE REFERENCE: 98-338
; CURRENT APPLICATION NUMBER: US/09/120,653D
; CURRENT FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: KR 98-10108
; PRIOR FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 4
; LENGTH: 2330
; TYPE: DNA
; ORGANISM: DIABETES-SPECIFIC RETROVIRUS ENDOGENOUS ERV-9
US-09-120-653D-4

Query Match      25.0%; Score 28.8; DB 3; Length 2330;
Best Local Similarity 62.5%; Pred. No. 3.5;
Matches 45; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 16 AGCATAAACTAGACACATCCTCATGCTTTTGAGGTCTAATCATTTGGATTGTTGCTTT 75
DB 179 AGCCTATAACCAAGTATTTCTCCCACTCTCTCAAGTTGTAATTTGGAGACTTTGCTCTTT 238

QY 76 CAGATGCTTTTC 87
DB 239 CACATGCTTTTC 250

RESULT 8
US-09-120-653D-1
; Sequence 1, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
; APPLICANT: YOON, JI-WON
; APPLICANT: JUN, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: AHN, JONG-SEONG
; APPLICANT: HA, YOUNG-JU
; APPLICANT: CHUNG, SOO-IL
; TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
; FILE REFERENCE: 98-338
; CURRENT APPLICATION NUMBER: US/09/120,653D
; CURRENT FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: KR 98-10108
; PRIOR FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 33
```

```
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 1
; LENGTH: 3910
; TYPE: DNA
; ORGANISM: DIABETES-SPECIFIC ENDOGENOUS RETROVIRUS ERV-9
US-09-120-653D-1

Query Match      25.0%; Score 28.4; DB 3; Length 3910;
Best Local Similarity 62.5%; Pred. No. 4.4;
Matches 45; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 16 AGCATAAACTAGACACATCTCTGCTTTTGAGGTCTAAATCATTTGGATTTGTCCTTT 75
Db 575 AGCCTATAACAGGATTTCTCCACCTCTCCAGTTGTAATGGAGACATTTGCTCTTTT 634
Qy 76 CAGATGGCTTTC 87
Db 635 CACATGCCCTTTC 646

RESULT 9
US-09-949-016-16928
; Sequence 16928, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16928
; LENGTH: 524032
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)...(524032)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16928

Query Match      24.7%; Score 28.4; DB 4; Length 524032;
Best Local Similarity 56.4%; Pred. No. 47;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 15 TAGCATAAACTAGACACATCTCTGCTTTTGAGGTCTAAATCATTTGATTTGTCCTT 74
Db 270680 TACTAAATAAGTAGTAACAAAAAATGATGTTTATTTTCATTCATTATATTGTGATTTT 270739
Qy 75 TCAGATGCTTTCCGAGAGATGAGATTCAT 108
Db 270740 CTAATGGGTGCTCTAGGAAGAAATGATACAT 270773

RESULT 10
US-09-949-016-16928
; Sequence 16928, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16928
; LENGTH: 524032
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)...(524032)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16928

Query Match      24.7%; Score 28.4; DB 4; Length 524032;
Best Local Similarity 56.4%; Pred. No. 47;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 15 TAGCATAAACTAGACACATCTCTGCTTTTGAGGTCTAAATCATTTGATTTGTCCTT 74
Db 270680 TACTAAATAAGTAGTAACAAAAAATGATGTTTATTTTCATTCATTATATTGTGATTTT 270739
Qy 75 TCAGATGCTTTCCGAGAGATGAGATTCAT 108
Db 270740 CTAATGGGTGCTCTAGGAAGAAATGATACAT 270773

RESULT 11
US-09-949-016-16930
; Sequence 16930, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16930
; LENGTH: 524032
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)...(524032)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16930

Query Match      24.7%; Score 28.4; DB 4; Length 524032;
Best Local Similarity 56.4%; Pred. No. 47;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 15 TAGCATAAACTAGACACATCTCTGCTTTTGAGGTCTAAATCATTTGATTTGTCCTT 74
Db 270680 TACTAAATAAGTAGTAACAAAAAATGATGTTTATTTTCATTCATTATATTGTGATTTT 270739
Qy 75 TCAGATGCTTTCCGAGAGATGAGATTCAT 108
Db 270740 CTAATGGGTGCTCTAGGAAGAAATGATACAT 270773
```

```
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16929
; LENGTH: 524032
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)...(524032)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16929

Query Match      24.7%; Score 28.4; DB 4; Length 524032;
Best Local Similarity 56.4%; Pred. No. 47;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 15 TAGCATAAACTAGACACATCTCTGCTTTTGAGGTCTAAATCATTTGATTTGTCCTT 74
Db 270680 TACTAAATAAGTAGTAACAAAAAATGATGTTTATTTTCATTCATTATATTGTGATTTT 270739
Qy 75 TCAGATGCTTTCCGAGAGATGAGATTCAT 108
Db 270740 CTAATGGGTGCTCTAGGAAGAAATGATACAT 270773
```

```
RESULT 11
US-09-949-016-16930
; Sequence 16930, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16930
; LENGTH: 524032
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)...(524032)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16930
```

```
Query Match      24.7%; Score 28.4; DB 4; Length 524032;
Best Local Similarity 56.4%; Pred. No. 47;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 15 TAGCATAAACTAGACACATCTCTGCTTTTGAGGTCTAAATCATTTGATTTGTCCTT 74
Db 270680 TACTAAATAAGTAGTAACAAAAAATGATGTTTATTTTCATTCATTATATTGTGATTTT 270739
Qy 75 TCAGATGCTTTCCGAGAGATGAGATTCAT 108
Db 270740 CTAATGGGTGCTCTAGGAAGAAATGATACAT 270773
```

```
RESULT 12
US-09-949-016-16931
; Sequence 16931, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16931
; LENGTH: 524032
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(524032)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16931

Query Match      24.7%; Score 28.4; DB 4; Length 524032;
Best Local Similarity 56.4%; Pred. No. 47;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy      15 TAGCATAAACTAGACACATCCTCATGCTTTTGGAGTCTTAATCATGATTGTTTCCTT 74
Db      270680 TACTAAATAAGTAGTAACAAAAAATGATGTTTATTTTCATTCAATATATTGTGATTTT 270739

Qy      75 TCAGATGCTTTCCGAGAGATGAGATTGATGAT 108
Db      270740 CTAATGGGTGCTCTAGGAGGAAAATGATACAT 270773

RESULT 13
US-09-949-016-14340
; Sequence 14340, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14340
; LENGTH: 529885
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(529885)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14340

Query Match      24.7%; Score 28.4; DB 4; Length 529885;
Best Local Similarity 56.4%; Pred. No. 48;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy      15 TAGCATAAACTAGACACATCCTCATGCTTTTGGAGTCTTAATCATGATTGTTTCCTT 74
Db      270680 TACTAAATAAGTAGTAACAAAAAATGATGTTTATTTTCATTCAATATATTGTGATTTT 270739

Qy      75 TCAGATGCTTTCCGAGAGATGAGATTGATGAT 108
Db      270740 CTAATGGGTGCTCTAGGAGGAAAATGATACAT 270773

RESULT 14
US-09-949-016-14341
; Sequence 14341, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14341
; LENGTH: 529885
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(529885)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14341

Query Match      24.7%; Score 28.4; DB 4; Length 529885;
Best Local Similarity 56.4%; Pred. No. 48;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy      15 TAGCATAAACTAGACACATCCTCATGCTTTTGGAGTCTTAATCATGATTGTTTCCTT 74
Db      270680 TACTAAATAAGTAGTAACAAAAAATGATGTTTATTTTCATTCAATATATTGTGATTTT 270739

Qy      75 TCAGATGCTTTCCGAGAGATGAGATTGATGAT 108
Db      270740 CTAATGGGTGCTCTAGGAGGAAAATGATACAT 270773

RESULT 15
US-09-949-016-14342
; Sequence 14342, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14342
; LENGTH: 529885
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(529885)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14342

Query Match      24.7%; Score 28.4; DB 4; Length 529885;
Best Local Similarity 56.4%; Pred. No. 48;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy      15 TAGCATAAACTAGACACATCCTCATGCTTTTGGAGTCTTAATCATGATTGTTTCCTT 74
Db      270680 TACTAAATAAGTAGTAACAAAAAATGATGTTTATTTTCATTCAATATATTGTGATTTT 270739

Qy      75 TCAGATGCTTTCCGAGAGATGAGATTGATGAT 108
Db      270740 CTAATGGGTGCTCTAGGAGGAAAATGATACAT 270773
```

```
; LENGTH: 529885
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(529885)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14342

Query Match      24.7%; Score 28.4; DB 4; Length 529885;
Best Local Similarity 56.4%; Pred. No. 48;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy      15 TAGCATAAACTAGACACATCCTCATGCTTTTGAGGCTCTAATCATTTGGATTTTGTCTT 74
Db      270680 TACTAAATAAGTAGTACAAAAAATGATGTTTATTTTCATTTCATTATATTGTGATTTT 270739

Qy      75 TCAGATGGCTTTCCCGAAGATGAGATTGATGTAT 108
Db      270740 CTAAATGGGTGTCCTTAGGAAGGAAATGATACAT 270773
```

Search completed: September 22, 2005, 17:02:04  
Job time : 31.9415 secs

**This Page Blank (uspro)**





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2005, 14:50:12 ; Search time 298.529 Seconds  
(without alignments)  
2575.655 Million cell updates/sec

Title: US-10-777-828-8\_COPY\_1\_115

Perfect score: 115

Sequence: 1 ccacctcagcctctagcat.....gagattgatgtatctgcc 115

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 7400704 seqs, 3343079526 residues

Total number of hits satisfying chosen parameters: 14801408

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/ECT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/ECTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*
- 19: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq.\*
- 20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq.\*
- 21: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq.\*
- 22: /cgn2\_6/ptodata/1/pubpna/US10J\_NEW\_PUB.seq.\*
- 23: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq.\*
- 24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 25: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 26: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115	100.0	1011	19	US-10-777-828-8
2	113.4	98.6	596	16	Sequence 8, Appli
c 3	33	28.7	78025	14	Sequence 13121, A
4	32	27.8	495269	17	US-10-020-141-9
c 5	32	27.8	3011208	17	Sequence 9, Appli
6	30.8	26.8	3796	19	US-10-398-221-8
7	30.6	26.6	2000	9	Sequence 8, Appli
					Sequence 2058, Ap
					Sequence 158, App
					Sequence 4201, Ap

c 9	30.6	26.6	2000	11	US-09-938-842A-4201	Sequence 4201, Ap
c 10	30.4	26.4	2000	11	US-09-938-842A-3143	Sequence 3143, Ap
c 11	30.4	26.4	2000	11	US-09-938-842A-3143	Sequence 3143, Ap
c 12	30.2	26.3	26027	13	US-10-496-011-4	Sequence 4, Appli
c 13	29.8	25.9	389	17	US-10-087-192-298	Sequence 298, App
c 14	29.8	25.9	389	18	US-10-242-535A-20929	Sequence 20929, A
c 15	29.8	25.9	389	18	US-10-085-783A-20929	Sequence 20929, A
c 16	29.8	25.9	4506	18	US-10-106-698-604	Sequence 604, App
c 17	29.8	25.9	4506	24	US-10-467-042-27	Sequence 27, Appli
c 18	29.4	25.6	568	22	US-11-046-868-27	Sequence 27, Appli
c 19	29.4	25.6	600	22	US-10-972-079-94291	Sequence 94291, A
c 20	29.4	25.6	1286	9	US-10-972-079-94290	Sequence 94290, A
c 21	29.4	25.6	1286	24	US-09-755-016-7	Sequence 7, Appli
c 22	29.4	25.6	1443	16	US-11-049-613-7	Sequence 7, Appli
c 23	29.4	25.6	103747	19	US-10-114-153-81	Sequence 81, Appli
c 24	29.2	25.4	487	13	US-10-450-826-46	Sequence 46, Appli
c 25	29.2	25.4	487	13	US-10-027-632-276926	Sequence 276926,
c 26	29.2	25.4	487	17	US-10-027-632-276927	Sequence 276927,
c 27	29.2	25.4	487	17	US-10-027-632-276927	Sequence 276927,
c 28	29.2	25.4	799	20	US-10-425-115-174734	Sequence 174734,
c 29	29.2	25.4	5683	15	US-10-311-455-1780	Sequence 1780, Ap
c 30	29.2	25.4	5683	15	US-10-240-485-144	Sequence 144, App
c 31	29.2	25.4	71843	22	US-10-488-292-3	Sequence 3, Appli
c 32	29	25.2	359	8	US-08-781-986A-2970	Sequence 2970, Ap
c 33	29	25.2	359	18	US-10-329-624-2970	Sequence 2970, Ap
c 34	29	25.2	1334	18	US-10-424-599-138167	Sequence 138167,
c 35	29	25.2	3251	13	US-10-027-632-115828	Sequence 115828,
c 36	29	25.2	3251	13	US-10-027-632-115829	Sequence 115829,
c 37	29	25.2	3251	13	US-10-027-632-115829	Sequence 115829,
c 38	29	25.2	3251	17	US-10-027-632-115828	Sequence 115828,
c 39	29	25.2	3251	17	US-10-027-632-115829	Sequence 115829,
c 40	29	25.2	3251	17	US-10-027-632-115830	Sequence 115830,
c 41	28.8	25.0	602	13	US-10-027-632-247342	Sequence 247342,
c 42	28.8	25.0	602	13	US-10-027-632-247343	Sequence 247343,
c 43	28.8	25.0	602	13	US-10-027-632-247344	Sequence 247344,
c 44	28.8	25.0	602	13	US-10-027-632-247345	Sequence 247345,
c 45	28.8	25.0	602	17	US-10-027-632-247342	Sequence 247342,

#### ALIGNMENTS

RESULT 1  
US-10-777-828-8  
; Sequence 8, Application US/10777828  
; Publication No. US20040142425A1  
; GENERAL INFORMATION:  
; APPLICANT: Clausen, Henrik  
; APPLICANT: Amado, Margarita  
; TITLE OF INVENTION: UDP-GALACTOSE: BETA-N-ACETYL-GLUCOSAMINE BETA 1,3  
; FILE REFERENCE: 7188-157  
; CURRENT APPLICATION NUMBER: US/10777,828  
; CURRENT FILING DATE: 2004-02-12  
; PRIOR APPLICATION NUMBER: US/09/831,630  
; PRIOR FILING DATE: 2001-05-10  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 1011  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (79)..(1008)  
US-10-777-828-8

Query Match 100.0%; Score 115; DB 19; Length 1011;  
Best Local Similarity 100.0%; Pred. No. 1.5e-28;  
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCACCTCAGCCTCCTAGCATAAACTAGACACATCCTCATGCTTTTGGAGTCTAATCATT 60

```
Db 1 CCACCTGAGCTCCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGTCCTAATCATT 60
Qy 61 GGATTTTGTCTCTTCAGATGGCTTCCGGAAGATGAGATTGATGATATCTGCC 115
Db 61 GGATTTTGTCTCTTCAGATGGCTTCCGGAAGATGAGATTGATGATATCTGCC 115

RESULT 2
US-10-029-386-13121
; Sequence 13121, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029.386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13121
; LENGTH: 596
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR21.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.71
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.53
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.59
; OTHER INFORMATION: NT HIT: g116170760, EVALUE 0.00e+00
; OTHER INFORMATION: EST HUMAN HIT: AJ003597.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q9Y2C3, EVALUE 3.00e-59
US-10-029-386-13121

Query Match 98.6%; Score 113.4; DB 16; Length 596;
Best Local Similarity 99.1%; Pred. No. 4.3e-28;
Matches 114; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCACCTGAGCTCCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGTCCTAATCATT 60
Db 187 CCACCTGAGCTCCTAGCATAAACTAGACACATCCTCATGCTTTTGAGGTCCTAATCATT 246

Qy 61 GGATTTTGTCTCTTCAGATGGCTTCCGGAAGATGAGATTGATGATATCTGCC 115
Db 247 GGATTTTGTCTCTTCAGATGGCTTCCGGAAGATGAGATTGATGATATCTGCC 301

RESULT 3
US-10-020-141-9/c
; Sequence 9, Application US/10020141
; Publication No. US2003092013A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Jeanette
; APPLICANT: Ableson, Allen
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE
; FILE REFERENCE: MMF-002
; CURRENT APPLICATION NUMBER: US/10/020.141
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/313,097
; PRIOR FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/327,485
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 78025
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-020-141-9
```

```
Query Match 28.7%; Score 33; DB 14; Length 78025;
Best Local Similarity 65.8%; Pred. No. 6.8;
Matches 48; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 15 TAGCATAAACTAGACACATCCTCATGCTTTTGAGGTCCTAATCATTGATTTTGTCTCTT 74
Db 5992 TAGCTGTAAACAGGTACTTCTCCACCTTCTCAGTTGTAAATGGGAGACTTTTGTCTCTT 5933

Qy 75 TCAGATGGCTTTC 87
Db 5932 TCACATGGCTTTC 5920

RESULT 4
US-10-398-221-8
; Sequence 8, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398.221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 495269
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-8

Query Match 27.8%; Score 32; DB 17; Length 495269;
Best Local Similarity 58.3%; Pred. No. 31;
Matches 56; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 16 AGCATAAACTAGACACATCCTCATGCTTTTGAGGTCCTAATCATTGATTTTGTCTCTT 75
Db 390225 AGCATACACGACTCATAGAGAATGCACCTGCTTGTCTAATGGTTGATATTATTGATA 390284

Qy 76 CAGATGGCTTTCCGGAAGATGAGATTGATGATATC 111
Db 390285 CGCCTGTATGCGTGAATTCGGATTGGTTTAAATC 390320

RESULT 5
US-10-398-221-2058/c
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398.221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2058
; LENGTH: 3011208
; TYPE: DNA
```

```

; ORGANISM: Listeria innocua
US-09-938-842A-4201
Query Match      26.6%; Score 30.6; DB 9; Length 2000;
Best Local Similarity 55.0%; Pred. No. 10;
Matches 60; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 2 CACCTCAGCCTCTAGCATAAAGCTAGACACATCTCTCATGCTTTTGGAGGTCTAAATCATTTG 61
Db 1077 CACCAAAATAAATCAAGCAGCGAACTCGAAAGATGATGATGATGATGATGATGATGATG 1136

QY 62 GATTTTGTTCTTTCAGATGCTTCCCGAAGATGAGATGATGATGATGATGATGATGATGAT 110
Db 1137 TGGATTGAATCATTTGGATTGGTTACCAAAAAAGAGAGATAATTT 1185

RESULT 8
US-09-938-842A-4201
; Sequence 4201, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4201
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4201
Query Match      26.6%; Score 30.6; DB 11; Length 2000;
Best Local Similarity 55.0%; Pred. No. 10;
Matches 60; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 2 CACCTCAGCCTCTAGCATAAAGCTAGACACATCTCTCATGCTTTTGGAGGTCTAAATCATTTG 61
Db 1077 CACCAAAATAAATCAAGCAGCGAACTCGAAAGATGATGATGATGATGATGATGATGATG 1136

QY 62 GATTTTGTTCTTTCAGATGCTTCCCGAAGATGAGATGATGATGATGATGATGATGATGAT 110
Db 1137 TGGATTGAATCATTTGGATTGGTTACCAAAAAAGAGAGATAATTT 1185

RESULT 9
US-09-938-842A-3143/c
; Sequence 3143, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4201
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4201
Query Match      26.8%; Score 30.8; DB 19; Length 3796;
Best Local Similarity 61.0%; Pred. No. 11;
Matches 50; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 31 ACATCCTCATGCTTTTGGAGTCTAATCATGATTTGTTCTTTCAGATGCTTTTCCCG 90
Db 3311 AAATGTAAGTATATGATTTATCTAGGATTTTTTTTTTAAAGATTTTAAATG 3370

QY 91 AAGATGAGATGATGATATATCT 112
Db 3371 GAGGTGGATTGATTTAAATAT 3392

RESULT 7
US-09-938-842A-4201
; Sequence 4201, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4201
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-433-793-158
; Sequence 158, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/433,793
; CURRENT FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 158
; LENGTH: 3796
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-158
Query Match      26.8%; Score 30.8; DB 19; Length 3796;
Best Local Similarity 61.0%; Pred. No. 11;
Matches 50; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 31 ACATCCTCATGCTTTTGGAGTCTAATCATGATTTGTTCTTTCAGATGCTTTTCCCG 90
Db 3311 AAATGTAAGTATATGATTTATCTAGGATTTTTTTTTTAAAGATTTTAAATG 3370

QY 91 AAGATGAGATGATGATATATCT 112
Db 3371 GAGGTGGATTGATTTAAATAT 3392

```

```

; ORGANISM: Listeria innocua
US-09-938-842A-4201
Query Match      26.6%; Score 30.6; DB 9; Length 2000;
Best Local Similarity 55.0%; Pred. No. 10;
Matches 60; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 2 CACCTCAGCCTCTAGCATAAAGCTAGACACATCTCTCATGCTTTTGGAGGTCTAAATCATTTG 61
Db 1077 CACCAAAATAAATCAAGCAGCGAACTCGAAAGATGATGATGATGATGATGATGATGATG 1136

QY 62 GATTTTGTTCTTTCAGATGCTTCCCGAAGATGAGATGATGATGATGATGATGATGATGAT 110
Db 1137 TGGATTGAATCATTTGGATTGGTTACCAAAAAAGAGAGATAATTT 1185

RESULT 8
US-09-938-842A-4201
; Sequence 4201, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4201
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4201
Query Match      26.6%; Score 30.6; DB 11; Length 2000;
Best Local Similarity 55.0%; Pred. No. 10;
Matches 60; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 2 CACCTCAGCCTCTAGCATAAAGCTAGACACATCTCTCATGCTTTTGGAGGTCTAAATCATTTG 61
Db 1077 CACCAAAATAAATCAAGCAGCGAACTCGAAAGATGATGATGATGATGATGATGATGATG 1136

QY 62 GATTTTGTTCTTTCAGATGCTTCCCGAAGATGAGATGATGATGATGATGATGATGATGAT 110
Db 1137 TGGATTGAATCATTTGGATTGGTTACCAAAAAAGAGAGATAATTT 1185

RESULT 9
US-09-938-842A-3143/c
; Sequence 3143, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4201
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4201
Query Match      26.8%; Score 30.8; DB 19; Length 3796;
Best Local Similarity 61.0%; Pred. No. 11;
Matches 50; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 31 ACATCCTCATGCTTTTGGAGTCTAATCATGATTTGTTCTTTCAGATGCTTTTCCCG 90
Db 3311 AAATGTAAGTATATGATTTATCTAGGATTTTTTTTTTAAAGATTTTAAATG 3370

QY 91 AAGATGAGATGATGATATATCT 112
Db 3371 GAGGTGGATTGATTTAAATAT 3392

RESULT 7
US-09-938-842A-4201
; Sequence 4201, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4201
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-433-793-158
; Sequence 158, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/433,793
; CURRENT FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 158
; LENGTH: 3796
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-158
Query Match      26.8%; Score 30.8; DB 19; Length 3796;
Best Local Similarity 61.0%; Pred. No. 11;
Matches 50; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 31 ACATCCTCATGCTTTTGGAGTCTAATCATGATTTGTTCTTTCAGATGCTTTTCCCG 90
Db 3311 AAATGTAAGTATATGATTTATCTAGGATTTTTTTTTTAAAGATTTTAAATG 3370

QY 91 AAGATGAGATGATGATATATCT 112
Db 3371 GAGGTGGATTGATTTAAATAT 3392

```

```
; TITLE OF INVENTION: POLYNUCLEOTIDE AND PROTEIN INVOLVED IN SYNAPTOGENESIS, VARIANTS
; TITLE OF INVENTION: THEREOF, AND THEIR THERAPEUTIC AND DIAGNOSTIC USES
; FILE REFERENCE: 253820USXPCT
; CURRENT APPLICATION NUMBER: US/10/496,011
; CURRENT FILING DATE: 2004-05-28
; PRIOR APPLICATION NUMBER: PCT/FR02/04134
; PRIOR FILING DATE: 2002-11-28
; PRIOR APPLICATION NUMBER: CA2364106
; PRIOR FILING DATE: 2001-11-30
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 4
; LENGTH: 335199
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (10156)..(10298)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (108500)..(109001)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (205869)..(205928)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (209526)..(209679)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (235028)..(235139)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (238090)..(238212)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (310597)..(310783)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (316139)..(316929)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (326822)..(330136)
; US-10-496-011-4
;
Query Match 26.4%; Score 30.4; DB 21; Length 335199;
Best Local Similarity 63.9%; Pred. No. 93;
Matches 46; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 16 AGCATAAACTAGACACATCTCATGCTTTTGGAGTCTAATCATTTGATTTTGTCTTTT 75
Db 304789 AGCTGTAAACAGGATTTCTCCAGCTCTCAGTTGTAATGGGAGACTTTGCTCTTTT 304730
QY 76 CAGATGCTTTTC 87
Db 304729 CACATGCTTTTC 304718

RESULT 12
US-10-087-192-298/c
; Sequence 298, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
;

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 8143
; LENGTH: 2090
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-3143
;
Query Match 26.4%; Score 30.4; DB 11; Length 2000;
Best Local Similarity 59.1%; Pred. No. 12;
Matches 52; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 19 ATAAACTAGACACATCTCATGCTTTTGGAGTCTAATCATTTGATTTTGTCTTTTTCAG 78
Db 203 ATTTATCTGCTAATCTCATTTTACCTGTTGATCTTGGGTTTGCTTCTTTAT 144
QY 79 ATGGCTTTCCCGAAGATGAGATTGATGT 106
Db 143 GTCGGCTTTGGGAGATTGGTTGATTT 116

RESULT 10
US-09-938-842A-3143/c
; Sequence 3143, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kröps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 8143
; LENGTH: 2090
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-3143
;
Query Match 26.4%; Score 30.4; DB 11; Length 2000;
Best Local Similarity 59.1%; Pred. No. 12;
Matches 52; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 19 ATAAACTAGACACATCTCATGCTTTTGGAGTCTAATCATTTGATTTTGTCTTTTTCAG 78
Db 203 ATTTATCTGCTAATCTCATTTTACCTGTTGATCTTGGGTTTGCTTCTTTAT 144
QY 79 ATGGCTTTCCCGAAGATGAGATTGATGT 106
Db 143 GTCGGCTTTGGGAGATTGGTTGATTT 116

RESULT 11
US-10-496-011-4/c
; Sequence 4, Application US/10496011
; Publication No. US20050118588A1
; GENERAL INFORMATION:
; APPLICANT: BOURGERON, THOMAS
; APPLICANT: JANAIN, STEPHANE
; APPLICANT: QUACH, HELENE
; APPLICANT: BETANCUR, CATALINA
; APPLICANT: LEBROYER, MARION
; APPLICANT: GILLBERG, CHRISTOPHER
```

NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 298  
; LENGTH: 260027  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(260027)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-087-192-298

Query Match 26.3%; Score 30.2; DB 13; Length 260027;  
Best Local Similarity 62.7%; Pred. No. 99;  
Matches 47; Conservative 0; Mismatches 28; Indels 0; Gaps 0;  
QY 16 AGCAATAAAGTAGACACATCCTCATGCTTTTGAGGCTCTAATCATGATGGATTTGTTCTCTTT 75  
Db 181901 AGCTGTAAACAGGTATTTCCTCCACCTCCTGAGTTGTAATGGAGACTTTGCTCTTTT 181842  
QY 76 CAGATGGCTTTCCCG 90  
Db 181841 CACATGCTTTCTCG 181827

RESULT 13  
US-10-242-535A-20929  
; Sequence 20929, Application US/10242535A  
; Publication No. US20040013663A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2005  
; CURRENT APPLICATION NUMBER: US/10/242,535A  
; CURRENT FILING DATE: 2002-09-12  
; PRIOR APPLICATION NUMBER: US 10/085,783  
; PRIOR FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 20929  
; LENGTH: 389  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (91)..(91)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-242-535A-20929

Query Match 25.9%; Score 29.8; DB 17; Length 389;  
Best Local Similarity 59.8%; Pred. No. 9.8;  
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 34 TCCTCATGCTTTTGAGGCTCTAATCATGATGGATTTGTTCTCTTCAGATGGCTTTCCCGAAG 93  
Db 250 TACTGAAGTTTAAAAAATTAATCATTTTCTCTGTTCACTTTACATGTTTTTCTGGTGG 309  
QY 94 ATGAGATTGATGTATATCTGCC 115  
Db 310 NGTAAATGAAGTATGCTGCC 331

US-10-242-535A-20929  
; Sequence 20929, Application US/10242535A  
; Publication No. US20040013663A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2005  
; CURRENT APPLICATION NUMBER: US/10/242,535A  
; CURRENT FILING DATE: 2002-09-12  
; PRIOR APPLICATION NUMBER: US 10/085,783  
; PRIOR FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 20929  
; LENGTH: 389  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (91)..(91)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-242-535A-20929

Query Match 25.9%; Score 29.8; DB 17; Length 389;  
Best Local Similarity 59.8%; Pred. No. 9.8;  
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 34 TCCTCATGCTTTTGAGGCTCTAATCATGATGGATTTGTTCTCTTCAGATGGCTTTCCCGAAG 93  
Db 250 TACTGAAGTTTAAAAAATTAATCATTTTCTCTGTTCACTTTACATGTTTTTCTGGTGG 309  
QY 94 ATGAGATTGATGTATATCTGCC 115  
Db 310 NGTAAATGAAGTATGCTGCC 331

RESULT 14  
US-10-085-783A-20929  
; Sequence 20929, Application US/10085783A  
; Publication No. US20040037841A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2002  
; CURRENT APPLICATION NUMBER: US/10/085,783A  
; CURRENT FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 20929  
; LENGTH: 389  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (91)..(91)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-085-783A-20929

Query Match 25.9%; Score 29.8; DB 18; Length 389;  
Best Local Similarity 59.8%; Pred. No. 9.8;  
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 34 TCCTCATGCTTTTGAGGCTCTAATCATGATGGATTTGTTCTCTTCAGATGGCTTTCCCGAAG 93  
Db 250 TACTGAAGTTTAAAAAATTAATCATTTTCTCTGTTCACTTTACATGTTTTTCTGGTGG 309  
QY 94 ATGAGATTGATGTATATCTGCC 115  
Db 310 NGTAAATGAAGTATGCTGCC 331

RESULT 15  
US-10-106-698-604  
; Sequence 604, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides  
; FILE REFERENCE: PA005P1  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03  
; NUMBER OF SEQ ID NOS: 8564  
; SOFTWARE: PatentIn Ver. 3.0  
; SEQ ID NO 604  
; LENGTH: 682  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (673)..(673)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-10-106-698-604

Query Match 25.9%; Score 29.8; DB 15; Length 682;  
 Best Local Similarity 60.5%; Pred. No. 12;  
 Matches 49; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
 Qy 28 GACACATCCTCATGCTTTTGAGGTCTAATCATTTGGATTGTTCTCTTTTCAGATGGCTTTC 87  
 Db 456 GAGCTCCCTCATGCTTTTAAGGTAGCATCATTTGGATGCTTTGGCTTTTGGATTTT 515  
 Qy 88 CCGAAGATGAGATTGATGTAT 108  
 Db 516 CTGAACACAGCTAATGTTGTGT 536

Search completed: September 22, 2005, 22:20:24  
 Job time : 303.529 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: September 22, 2005, 07:38:28 ; Search time 2670.69 Seconds  
(without alignments)  
10595.706 Million cell updates/sec

Title: US-10-777-828-8\_COPY\_428\_1011

Perfect score: 584

Sequence: 1 gggacattatccgaagat.....aagattgcgcctgtctga 584

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_on.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	584	100.0	933	9	AF145784	Homo sapi
2	584	100.0	1011	6	BD228354	UDP-galac
3	584	100.0	2762	9	AB020337	Homo sapi
4	584	100.0	2775	6	E38419	Novel poly
5	584	100.0	2921	9	AY372061	Homo sapi
6	584	100.0	10562	6	E38420	Novel poly
7	584	100.0	170121	9	AF064860	Homo sapi
8	584	100.0	340000	9	HS21C080	Homo sapi
9	583	99.8	933	9	HSA6078	Homo sapi
10	582.4	99.7	2494	6	CQ731786	Sequence
11	569.6	97.5	192219	9	RF43002119	Pan trogl
12	553	94.7	1576	9	AB041416	Homo sapi
13	541.2	92.7	1570	9	AB041415	Pan panis
14	532.6	91.2	1360	9	AB041412	Gorilla g
15	531	90.9	1565	9	AB041413	Homo sapi
16	530.8	90.9	937	9	AY231145	Macaca mu
17	530.2	90.8	1566	9	AB041414	Pan trogl
18	530	90.8	1579	9	AB041417	Pongo pyg
19	387.2	66.3	170108	2	AC150794	Bos tauru

20	368	63.0	927	10	AF254738	Mus muscu	
21	368	63.0	4933	10	BC057887	Mus muscu	
22	368	63.0	5069	10	BC051669	Mus muscu	
23	368	63.0	196900	2	AC020851	Mus muscu	
24	363.2	62.2	149964	2	AC120145	Mus muscu	
25	363.2	62.2	186956	2	AC120346	Mus muscu	
C	26	321.8	55.1	65400	2	AC109264	Mus muscu
C	27	213.8	36.6	97702	2	AC151319	Xenopus t
C	28	159.8	27.4	226720	2	BCX936311	Danio rer
29	158.2	27.1	2317	5	BC066477	Danio rer	
30	132.6	22.7	1950	5	BC082704	Xenopus l	
31	124.6	21.3	1474	5	BC075347	Xenopus t	
32	124.4	21.3	917	10	AB039137	Mus muscu	
33	124.4	21.3	917	10	AB039140	Mus muscu	
34	124.4	21.3	917	10	AB039141	Mus muscu	
35	124.4	21.3	917	10	AB039142	Mus muscu	
C	36	123.4	21.1	110000	2	AC096079_1	Continuation (2 of
C	37	123.4	21.1	231235	2	AC136678	Rattus no
C	38	123.4	21.1	231471	2	AC137010	Rattus no
C	39	123.4	21.1	244843	2	AC107162	Rattus no
40	123.4	21.1	264119	2	AC134162	Rattus no	
41	123	21.1	2049	5	BC080111	Xenopus l	
42	122.8	21.0	917	10	AB039134	Mus muscu	
43	122.8	21.0	917	10	AB039135	Mus muscu	
44	122.8	21.0	917	10	AB039136	Mus muscu	
45	122.8	21.0	917	10	AB039138	Mus muscu	

ALIGNMENTS

RESULT 1  
AF145784  
LOCUS  
DEFINITION Homo sapiens beta1.3 galactosyltransferase-V (B3GALT5) gene,  
complete cds.  
ACCESSION AF145784  
VERSION AF145784.1 GI:6409192  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Zhou,D., Berger,G.G. and Hennet,T.  
Molecular cloning of a human UDP-galactose-4-epimerase gene encoding an O-linked  
beta1,3 galactosyltransferase gene  
JOURNAL Eur. J. Biochem. 263 (2), 571-576 (1999)  
PUBMED 99337698  
REFERENCE 2 (bases 1 to 933)  
AUTHORS Zhou,D. and Hennet,T.  
TITLE Direct Submission  
Submitted (26-APR-1999) Physiology, University of Zurich,  
Winterthurerstrasse 190, Zurich 8057, Switzerland  
JOURNAL Location/Qualifiers  
FEATURES  
source  
1..933  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22"  
1..933  
/gene="B3GALT5"  
1..933  
/gene="B3GALT5"  
/note="glycosyltransferase"  
/codon\_start=1  
/product="beta1,3 galactosyltransferase-V"  
/protein\_id="AA07880.1"  
/db\_xref="GI:6409193"  
/translation="MAFFKRLMYICLLVLGALCLYFSWYSLNPFKEQSFYVKDGNF

```

LKLPTDTCROTPEPPLVLLVTSKQLAERMAIROTWGERMVKQKLTFFLLGTSS
AAETKVEDQSRHGDIIQKDFLDVYVNLTLTKWMLGIEWHRFCPOAFVWKTSDFE
INVDYLTELLKKNRTRFTFTGFLKNEPFRQPSKFWKSEYPMWRIPYDFPCSGRG
YVSGDVASQVYNVSKSVYIKLSDVFGVGLCLERLIRLBELHSHQPTFFPGGURFSVC
LFRIVACHFIKPRTLDDYQWALENSRGEDCPVP"

ORIGIN
Query Match 100.0%; Score 584; DB 9; Length 933;
Best Local Similarity 100.0%; Pred. No. 6.6e-151;
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGACATTATCCAGAAGGATTTCCTAGACGCTCTATTACAATCTGACCCCTGAAGACCATGA 60
Db 350 GGGACATTATCCAGAAGGATTTCCTAGACGCTCTATTACAATCTGACCCCTGAAGACCATGA 409

Qy 61 TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGGCGCGTGTGTGATGAAACACAGACT 120
Db 410 TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGGCGCGTGTGTGATGAAACACAGACT 469

Qy 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAACAGAACAA 180
Db 470 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAACAGAACAA 529

Qy 181 CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCCATCAGGCAGGCATTTCAGCA 240
Db 530 CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCCATCAGGCAGGCATTTCAGCA 589

Qy 241 AGTGGTTTGTTCAGTAAATCTGAATATCCGTGGGACAGGTACCCCAATTTCTGCTCCGGCA 300
Db 590 AGTGGTTTGTTCAGTAAATCTGAATATCCGTGGGACAGGTACCCCAATTTCTGCTCCGGCA 649

Qy 301 CCGGCTACGTTTCTGGGACAGCTGGCGAGTCAGGTGTACAATGTCTCAGAGCGTCC 360
Db 650 CCGGCTACGTTTCTGGGACAGCTGGCGAGTCAGGTGTACAATGTCTCAGAGCGTCC 709

Qy 361 CATACATTAAACTTGAAGACGCTGTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGAT 420
Db 710 CATACATTAAACTTGAAGACGCTGTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGAT 769

Qy 421 TGGAGGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC 480
Db 770 TGGAGGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC 829

Qy 481 TCTTTCAGGAGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGC 540
Db 830 TCTTTCAGGAGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGC 889

Qy 541 AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCCGCCCTGTCTGA 584
Db 890 AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCCGCCCTGTCTGA 933

RESULT 2
BD228354 1011 bp DNA linear PAT 17-JUL-2003
LOCUS
DEFINITION UDP-galactose: beta-N-acetyl-glucosamine beta 1,3
galactosyltransferases, beta 3 Gal-T5.
BD228354
BD228354.1 GI:33038124
VERSION
KEYWORDS JP 2002530071-A/8.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1. (bases 1 to 1011)
Clausen,H. and Amado,M.
JOURNAL UDP-galactose: beta-N-acetyl-glucosamine beta 1,3
galactosyltransferases, beta 3 Gal-T5
HENDRIK CLAUSEN
COMMENT Patent: JP 2002530071-A 8 17-SEP-2002;
OS Homo sapiens (human)
PN JP 2002530071-A/8
PD 17-SEP-2002
```

```

PF 11-NOV-1999 JP 2000582542
PR 13-NOV-1998 DK PA 199801483
PI HENDRIK CLAUSEN, MARGARIDA AMADO
PC C12N15/09, C12N1/15, C12N1/21, C12N5/10, C12N9/10, C12N15/
PC 00, C12N5/00
CC UDP-galactose: beta-N-acetyl-glucosamine beta 1,3 CC
galactosyltransferases,
CC beta 3 Gal-T5 Location/Qualifiers
FH Key Location/Qualifiers
FT CDS (79)..(1008).

FEATURES
source
1..1011
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 100.0%; Score 584; DB 6; Length 1011;
Best Local Similarity 100.0%; Pred. No. 6.6e-151;
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGACATTATCCAGAAGGATTTCCTAGACGCTCTATTACAATCTGACCCCTGAAGACCATGA 60
Db 428 GGGACATTATCCAGAAGGATTTCCTAGACGCTCTATTACAATCTGACCCCTGAAGACCATGA 487

Qy 61 TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGGCGCGTGTGTGATGAAACACAGACT 120
Db 488 TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGGCGCGTGTGTGATGAAACACAGACT 547

Qy 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAATGCTTCTGAAGAAAACAGAACAA 180
Db 548 CAGACATGTTTCATCAATGTTGACTATCTGACTGAATGCTTCTGAAGAAAACAGAACAA 607

Qy 181 CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCCATCAGGCAGGCATTTCAGCA 240
Db 608 CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCCATCAGGCAGGCATTTCAGCA 667

Qy 241 AGTGGTTTGTTCAGTAAATCTGAATATCCGTGGGACAGGTACCCCAATTTCTGCTCCGGCA 300
Db 668 AGTGGTTTGTTCAGTAAATCTGAATATCCGTGGGACAGGTACCCCAATTTCTGCTCCGGCA 727

Qy 301 CCGGCTACGTTTCTGGGACAGCTGGCGAGTCAGGTGTACAATGTCTCAGAGCGTCC 360
Db 728 CCGGCTACGTTTCTGGGACAGCTGGCGAGTCAGGTGTACAATGTCTCAGAGCGTCC 787

Qy 361 CATACATTAAACTTGAAGACGCTGTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGAT 420
Db 788 CATACATTAAACTTGAAGACGCTGTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGAT 847

Qy 421 TGGAGGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC 480
Db 848 TGGAGGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC 907

Qy 481 TCTTTCAGGAGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGC 540
Db 908 TCTTTCAGGAGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGC 967

Qy 541 AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCCGCCCTGTCTGA 584
Db 968 AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCCGCCCTGTCTGA 1011

RESULT 3
AB020337 2762 bp mRNA linear PRI 14-APR-2000
LOCUS
DEFINITION Homo sapiens mRNA for UDP-Gal:GlcNAc beta1,3-galactosyltransferase
5, complete cds.
AB020337
AB020337.1 GI:4835502
VERSION
KEYWORDS UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```



Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (sites)  
Ishihiki,S., Togayachi,A., Kudo,T., Nishihara,S., Watanabe,M.,  
Kubota,T., Kitajima,M., Shiraishi,N., Sasaki,K., Andoh,T. and  
Narimatsu,H.  
Cloning, expression, and characterization of a novel  
UDP-galactose:beta-N-acetylglucosamine  
betal,3-galactosyltransferase (beta3Gal-T5) responsible for  
synthesis of type 1 chain in colorectal and pancreatic epithelia  
and tumor cells derived therefrom  
J. Biol. Chem. 274 (18), 12499-12507 (1999)  
99230269  
10212226  
2 (bases 1 to 2762)  
Ishihiki,S., Togayachi,A. and Narimatsu,H.  
Direct Submission  
Submitted (20-NOV-1998) Hisashi Narimatsu, Soka University,  
Institute of Life Science; 1-236, Tangi-cho, Hachioji, Tokyo  
192-8577, Japan (E-mail:sishihiki@po.hijnet.or.jp,  
Tel:81-426-91-9466, Fax:81-426-91-9315)  
FEATURES  
source  
1. .2762  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.3"  
/cell\_line="Colo 205"  
/cell\_type="Adenocarcinoma"  
1. .2762  
/gene="betal,3-Galt 5"  
1. .273  
/gene="betal,3-Galt 5"  
/number=1  
274. .433  
/gene="betal,3-Galt 5"  
/number=3  
434. .2762  
/gene="betal,3-Galt 5"  
/number=4  
434. .1366  
/gene="betal,3-Galt 5"  
/codon\_start=1  
/product="UDP-Gal:GlcNAc betal,3-galactosyltransferase 5"  
/protein\_id="BAA77664.1"  
/db\_xref="GI:4835503"  
/translation="MAPPKRLMYICLLVLGALCLYFSWYSLNPPKQSFYVKDGNF  
LKLEPTDCRQTPPELVLLVTSRQLARMAIRQWGRNVRKQKLTFFLLGTTSS  
AAETKVDQESQRHDIQKDFLDVYVNLTKTWGIEWHRFCPOAFAVMKTDSDMF  
INVLYLTKLKKRTRFTFTGFLKLNFFIRQPFKSWFVSKSYPMWDRYPPFGSGTG  
YVFGSDVASQVYNSKSPYIKLEDFVGLCLRLNIRLELHSQPTFFPGGLRFSVC  
LFRIVACHFIKPRTLDDYQALENSRGEDCPVP"  
2762  
/gene="betal,3-Galt 5"  
/note="45 a nucleotides"  
polyA\_site  
ORIGIN  
Query Match 100.0%; Score 584; DB 9; Length 2762;  
Best Local Similarity 100.0%; Pred. No. 6.7e-151; Indels 0; Gaps 0;  
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GGGACATTATCCAGAAGGATTTCCTAGACGCTCTATTACAATCTGACCTGAAGACCATGA 60  
Db 783 GGGACATTATCCAGAAGGATTTCCTAGACGCTCTATTACAATCTGACCTGAAGACCATGA 842  
Qy 61 TGGGCATAGAAATGGGTCCATCGCTTTTCTCTCAGCGCGGCTTTGTGATGAAACAGACT 120  
Db 843 TGGGCATAGAAATGGGTCCATCGCTTTTCTCTCAGCGCGGCTTTGTGATGAAACAGACT 902  
Qy 121 CAGACATGTTTCATCAATCTGACTATCTGACTGCTTCTGAGGAAACAGACAA 180  
Db 903 CAGACATGTTTCATCAATCTGACTATCTGACTGCTTCTGAGGAAACAGACAA 962

Qy 181 CCAGGTTTTTCACTGGCTTCTTTGAAATCAATAGATTTCCCATAGGAGCCATTTCAGCA 240  
Db 963 CCAGGTTTTTCACTGGCTTCTTTGAAATCAATAGATTTCCCATAGGAGCCATTTCAGCA 1022  
Qy 241 AGTGGTTTGTGAGTAAATCTGAATATCGTGGGACAGGTACCCACCAATTCCTCGCGCA 300  
Db 1023 AGTGGTTTGTGAGTAAATCTGAATATCGTGGGACAGGTACCCACCAATTCCTCGCGCA 1082  
Qy 301 CCGGCTACGTGTTTCTCGCGACGCTGGCGAGTCAGGTGTACAAATGTCTCAGAGAGCGTCC 360  
Db 1083 CCGGCTACGTGTTTCTCGCGACGCTGGCGAGTCAGGTGTACAAATGTCTCAGAGAGCGTCC 1142  
Qy 361 CATACATTAAACTGGAAGACGCTGTTTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT 420  
Db 1143 CATACATTAAACTGGAAGACGCTGTTTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT 1202  
Qy 421 TGGAGGAGCTCCACTCCCGACGCGACCTTTTTCAGGGGGCTTAGCGTCTCCGATATGCC 480  
Db 1203 TGGAGGAGCTCCACTCCCGACGCGACCTTTTTCAGGGGGCTTAGCGTCTCCGATATGCC 1262  
Qy 481 TCTTCAGGAGGATCGTGGCTCGCCACTTTCATCAAGCCTCGGACTCTCTTGGACTACTGGC 540  
Db 1263 TCTTCAGGAGGATCGTGGCTCGCCACTTTCATCAAGCCTCGGACTCTCTTGGACTACTGGC 1322  
Qy 541 AGGCTCTAGAGAAATTCGCGGGGGGAAGATTGTCGCCCTGTCTGA 584  
Db 1323 AGGCTCTAGAGAAATTCGCGGGGGGAAGATTGTCGCCCTGTCTGA 1366  
RESULT 4  
E38419  
LOCUS E38419 Novel polypeptide. 2775 bp DNA linear PAT 31-JAN-2002  
DEFINITION E38419  
ACCESSION E38419  
VERSION E38419.1 GI:18626993  
KEYWORDS JP 2000245464-A/1.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2775)  
AUTHORS Narimatsu,H., Ishihiki,S., Togayachi,A. and Sasaki,K.  
TITLE Novel polypeptide  
JOURNAL Patent: JP 2000245464-A 1 12-SEP-2000;  
KYOWA HAKKO KOGYO CO LTD  
OS Homo sapiens (human)  
COMMENT PN JP 2000245464-A/1  
PD 12-SEP-2000  
PF 25-FEB-1999 JP 1999047571  
PI HISASHI NARIMATSU, SOICHIRO ISSHIKI, AKIRA TOGAYACHU, PI  
KATSUTOSHI SASAKI  
PC C12N15/09,A01K67/027,C12N1/21,C12N5/10,C12N9/10,C12P19/00, PC  
C12P21/02,  
PC C12P21/08,C12Q1/68,G01N33/53//C12N1/21,C12R1/185), (C12N5/10,  
PC C12R1/91),  
PC (C12P21/02,C12R1/185), (C12P21/02,C12R1/91), (C12N5/00,C12N5/00,  
PC (C12N5/00,C12R1/91)  
CC  
FH Key Location/Qualifiers  
FT CDS (402)..(1331).  
source  
1. .2775  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 100.0%; Score 584; DB 6; Length 2775;  
Best Local Similarity 100.0%; Pred. No. 6.7e-151; Indels 0; Gaps 0;  
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GGGACATTATCCAGAAGGATTTCCTAGACGCTCTATTACAATCTGACCTGAAGACCATGA 60

Db	751	GGGACATATCCAGAGATTTCTAGACGCTATTACATCTGACCCCTGAGACCATGA	810
Qy	61	TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGCGGCGTTTGTGATGAAACAGACT	120
Db	811	TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGCGGCGTTTGTGATGAAACAGACT	870
Qy	121	CAGACATGTTCAATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAAACAGAACAA	180
Db	871	CAGACATGTTCAATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAAACAGAACAA	930
Qy	181	CCAGGTTTTTCACTGGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGCGAGCCATTGACGA	240
Db	931	CCAGGTTTTTCACTGGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGCGAGCCATTGACGA	990
Qy	241	AGTGGTTTGTGACGTAATCTGAATATCCGTGGGACAGGTACCCACCATTTGCTCCCGGCA	300
Db	991	AGTGGTTTGTGACGTAATCTGAATATCCGTGGGACAGGTACCCACCATTTGCTCCCGGCA	1050
Qy	301	CCGCGTACGTTTCTTGGCGACGTTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT	360
Db	1051	CCGCGTACGTTTCTTGGCGACGTTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT	1110
Qy	361	CATACATTTAAACTGGAAGACGTTTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT	420
Db	1111	CATACATTTAAACTGGAAGACGTTTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT	1170
Qy	421	TGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC	480
Db	1171	TGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC	1230
Qy	481	TCCTTCAGAGAGATCGTGGCGCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTTGGC	540
Db	1231	TCCTTCAGAGAGATCGTGGCGCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTTGGC	1290
Qy	541	AGGCTCTAGAGAAATCCCGGGGGGAAGATTCTCGCGCTGTCTGA	584
Db	1291	AGGCTCTAGAGAAATCCCGGGGGGAAGATTCTCGCGCTGTCTGA	1334
RESULT 5	AY372061	2921 bp mRNA linear PRI 04-NOV-2003	
LOCUS	Homo sapiens beta-1,3-galactosyltransferase 5 (B3GALT5) mRNA,		
DEFINITION	complete cds.		
ACCESSION	AY372061		
VERSION	AY372061.1	GI:38045970	
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	Dunn, C.A., Medstrand, P. and Mager, D.L.		
JOURNAL	An endogenous retroviral long terminal repeat is the dominant		
PUBMED	promoter for human [beta]1,3-galactosyltransferase 5 in the colon		
REFERENCE	Proc. Natl. Acad. Sci. U.S.A. 100 (22), 12841-12846 (2003)		
AUTHORS	1. .2921		
TITLE	Submitted (20-AUG-2003) Terry Fox Laboratory, BC Cancer Agency, 601		
JOURNAL	West 10th Avenue, Vancouver, BC V5Z 1L3, Canada		
FEATURES	Location/Qualifiers		
source	1. .2921		
gene	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/chromosome="21"		
	/map="21q22.2"		
	1. .2921		
	/gene="B3GALT5"		
	/note="Beta3Gal-T5"		

CDS	593..1525		
	/gene="B3GALT5"		
	/note="beta3Gal-T5"		
	/codon_start=1		
	/product="beta-1,3-galactosyltransferase 5"		
	/protein_id="AAR08910.1"		
	/db_xref="GI:38045971"		
	/translation="MAPPPKMLMVI CLVLGALCLYFMSYSLNPFKEQSFVYKDGNF		
	LKLPDDRCRQTPFPLVLVLTSSHKQLAERMAIRQTWGERMVGKQLKTFLLGTTSS		
	AASKEVDQESQRHGD1 IQKDFLDVYNNLTMMGLVWHRFCPQAAFPVMTKDSDF		
	INDYITELLKKNNRTFRFTFKLNEPFIQRFPSKFWFSKSEYPMWRDYPFSCGTF		
	YVPSGVASQVNVNKSVPYIKLEDVFGVGLCLRLNIRLELHSQPTFFPGLRFSVC		
	LPRIIVACHFIKPTLLDYWQALNSRGEDCPV"		
ORIGIN			
	Query Match 100.0%; Score 584; DB 9; Length 2921;		
	Best Local Similarity 100.0%; Pred. No. 6.7e-151;		
	Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1	GGGACATATCCAGAGATTTCTAGACGCTATTACATCTGACCCCTGAGACCATGA	60
Db	942	GGGACATATCCAGAGAGATTTCTAGACGCTATTACATCTGACCCCTGAGACCATGA	1001
Qy	61	TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGCGGCGTTTGTGATGAAACAGACT	120
Db	1002	TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGCGGCGTTTGTGATGAAACAGACT	1061
Qy	121	CAGACATGTTCAATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAAACAGAACAA	180
Db	1062	CAGACATGTTCAATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAAACAGAACAA	1121
Qy	181	CCAGGTTTTTCACTGGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGCGAGCCATTGACGA	240
Db	1122	CCAGGTTTTTCACTGGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGCGAGCCATTGACGA	1181
Qy	241	AGTGGTTTGTGACGTAATCTGAATATCCGTGGGACAGGTACCCACCATTTGCTCCCGGCA	300
Db	1182	AGTGGTTTGTGACGTAATCTGAATATCCGTGGGACAGGTACCCACCATTTGCTCCCGGCA	1241
Qy	301	CCGCGTACGTTTCTTGGCGACGTTGGCGAGTCAAGTGTACAATGTCTCCAAGAGCGTCC	360
Db	1242	CCGCGTACGTTTCTTGGCGACGTTGGCGAGTCAAGTGTACAATGTCTCCAAGAGCGTCC	1301
Qy	361	CATACATTAACCTGGAAGAGTGTGTGGGGCTCTCGCTCGAAAGGCTGAACATCAGAT	420
Db	1302	CATACATTAACCTGGAAGAGTGTGTGGGGCTCTCGCTCGAAAGGCTGAACATCAGAT	1361
Qy	421	TGGAGGAGCTCCACTCCCGACGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC	480
Db	1362	TGGAGGAGCTCCACTCCCGACGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC	1421
Qy	481	TCCTTCAGAGAGATCGTGGCGCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTTGGC	540
Db	1422	TCCTTCAGAGAGATCGTGGCGCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTTGGC	1481
Qy	541	AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCCGCGCTGTCTGA	584
Db	1482	AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCCGCGCTGTCTGA	1525
RESULT 6	E38420	10562 bp DNA linear PAT 31-JAN-2002	
LOCUS	E38420		
DEFINITION	Novel polypeptide.		
ACCESSION	E38420		
VERSION	E38420.1	GI:18626994	
KEYWORDS	JP 2000245464-A/2.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	1 (bases 1 to 10562)		
	Narimatsu, H., Isshiki, S., Togayauchi, A. and Sasaki, K.		

TITLE Novel polypeptide  
JOURNAL Patent: JP 2000245464-A 2 12-SEP-2000;  
KYOWA HAKKO KOGYO CO LTD  
COMMENT OS Homo sapiens (human)  
PN JP 2000245464-A/2  
PD 12-SEP-2000  
PF 25-FEB-1999 JP 1999047571  
PR  
PI HISASHI NARIMATSU, SOICHIRO ISSHIKI, AKIRA TOGAYAUCHI, PI  
KATSUTOSHI SASAKI  
PC C12N15/09, A01K67/027, C12N1/21, C12N5/10, C12N9/10, C12P19/00, PC  
C12P21/02,  
PC C12P21/08, C12Q3/68, G01N33/53//C12N1/21, C12R1:185), (C12N5/10,  
PC C12R1:91),  
PC (C12P21/02, C12R1:185), (C12P21/02, C12R1:91), C12N15/00, C12N5/00,  
PC (C12N5/00, C12R1:91)  
CC  
Key Location/Qualifiers  
FH promoter (1)..(5000)  
FT exon (5001)..(5140)  
FT exon (5001)..(5273)  
FT exon (5459)..(5567)  
FT exon (7427)..(7586)  
FT exon (8234)..(10562).  
FEATURES source  
1..10562  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 100.0%; Score 584; DB 6; Length 10562;  
Best Local Similarity 100.0%; Pred. No. 6.9e-151;  
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GGGACATTATCCAGAGGATTCTCTAGACGCTATTACAACTGACCTCGAGGACCATGA 60  
Db 8583 GGGACATTATCCAGAGGATTCTCTAGACGCTATTACAACTGACCTCGAGGACCATGA 8642  
Qy 61 TGGGCATAGATGGTCCATCGCTTTCTCTCAGCGCGCTTTGTGATGAACACAGACT 120  
Db 8643 TGGGCATAGATGGTCCATCGCTTTCTCTCAGCGCGCTTTGTGATGAACACAGACT 8702  
Qy 121 CAGACATGTTTCATCAATGTTGACTATCTGACTATCTGACTGCTTTCTGAGAAACAGAACAA 180  
Db 8703 CAGACATGTTTCATCAATGTTGACTATCTGACTGCTTTCTGAGAAACAGAACAA 8762  
Qy 181 CCAGGTTTTTCATCGGCTTTCTGAAACTCAATGAGTTTCCATCAGGAGCCATTCAGCA 240  
Db 8763 CCAGGTTTTTCATCGGCTTTCTGAAACTCAATGAGTTTCCATCAGGAGCCATTCAGCA 8822  
Qy 241 AGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCA 300  
Db 8823 AGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCA 8882  
Qy 301 CCGGCTAGCTGTTTCTCGGACGCTGGGAGTCTGAGTGTACAAATGTTCTCCAGAGCGTCC 360  
Db 8883 CCGGCTAGCTGTTTCTCGGACGCTGGGAGTCTGAGTGTACAAATGTTCTCCAGAGCGTCC 8942  
Qy 361 CATACATTAAACTGGAAGACGTTGTTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT 420  
Db 8943 CATACATTAAACTGGAAGACGTTGTTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT 9002  
Qy 421 TGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTCTCCGATGCC 480  
Db 9003 TGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTCTCCGATGCC 9062  
Qy 481 TCTTCAGAGGATCGTGGCCTGCCACTTTCATCAGGCTCGGACTCTCTTGGACTACTGGC 540  
Db 9063 TCTTCAGAGGATCGTGGCCTGCCACTTTCATCAGGCTCGGACTCTCTTGGACTACTGGC 9122  
Qy 541 AGGCTCTAGAGATTCCCGGGGGGAAGATTGTCGCGCTGTCTGA 584

Db 9123 AGGCTCTAGAGATTCCCGGGGGGAAGATTGTCGCGCTGTCTGA 9166  
RESULT 7  
AF064860  
LOCUS Homo sapiens chromosome 21 clone PAC 70124 map 21q22.3, complete  
DEFINITION sequence.  
ACCESSION AF064860.2 GI:18958624  
VERSION AF064860  
KEYWORDS HTG; HTGS DRAFT  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 170121)  
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,  
Park,H.-S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.-K., Soeda,E.,  
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,  
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,  
Zimmermann,W., Reichwald,K., Rump,A., Schillhabel,M.B., Schudy,A.,  
Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,  
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordtsiek,G.,  
Hornischer,K., Brandt,P., Sharfe,M., Schoen,O., Desario,A.,  
Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,  
Hennig,S., Riesselmann,L., Dagand,E., Haaf,T., Wehrmeyer,S.,  
Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H.,  
Reinhardt,R. and Yaspo,M.Laure.  
The DNA sequence of human chromosome 21  
Nature 405 (6784), 311-319 (2000)  
REFERENCE 2 (bases 1 to 170121)  
AUTHORS Rump,A., Dagand,E., Hildmann,T., Nordtsiek,G., Drescher,B.,  
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.  
Direct Submission  
Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany  
REFERENCE 3 (bases 1 to 170121)  
AUTHORS Rump,A., Dagand,E., Hildmann,T., Nordtsiek,G., Drescher,B.,  
Weber,J., Schattevoy,R., Yaspo,M.-L., Rosenthal,A., Yaspo,M.-L. and  
Rosenthal,A.  
Direct Submission  
Submitted (27-PB-2002) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany  
REFERENCE 4 (bases 1 to 170121)  
AUTHORS Rump,A., Dagand,E., Hildmann,T., Nordtsiek,G., Drescher,B.,  
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.  
Direct Submission  
Submitted (05-MAR-2002) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany  
COMMENT On Feb 27, 2002 this sequence version replaced gi:3171153.  
FEATURES source  
1..170121  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.3"  
/clone="PAC 70124"  
ORIGIN  
Query Match 100.0%; Score 584; DB 9; Length 170121;  
Best Local Similarity 100.0%; Pred. No. 7.2e-151;  
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GGGACATTATCCAGAGGATTCTCTAGACGCTATTACAACTGACCTCGAGGACCATGA 60  
Db 90397 GGGACATTATCCAGAGGATTCTCTAGACGCTATTACAACTGACCTCGAGGACCATGA 90456  
Qy 61 TGGGCATAGATTGGTCCATCGCTTTTGTCTCAGCGCGCTTTTGTGATGAACACAGACT 120

```
Db 90457 TGGCATAGAAATGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAACACAGACT 90516
Qy 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGACTGACTGCTTCTGAAGAAAACAGAACAA 180
Db 90517 CAGACATGTTTCATCAATGTTGACTATCTGACTGACTGACTGCTTCTGAAGAAAACAGAACAA 90576
Qy 181 CCAGGTTTTTCACTGGGCTTCTTGAACATCAATGAGTTCCTCCATCAGGCGAGCCATTTCAGCA 240
Db 90577 CCAGGTTTTTCACTGGGCTTCTTGAACATCAATGAGTTCCTCCATCAGGCGAGCCATTTCAGCA 90636
Qy 241 AGTGGTTTGTGAGTAAATCTGAATATCCGTGGGAGAGTACCCACCAATTCGTCTCCGGCA 300
Db 90637 AGTGGTTTGTGAGTAAATCTGAATATCCGTGGGAGAGTACCCACCAATTCGTCTCCGGCA 90696
Qy 301 CCGGCTACGCTGTTTTCTGGCGAGCTGGCGAGTCAAGTGTACAATGTTCTCCAAGAGCGTCC 360
Db 90697 CCGGCTACGCTGTTTTCTGGCGAGCTGGCGAGTCAAGTGTACAATGTTCTCCAAGAGCGTCC 90756
Qy 361 CATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGAT 420
Db 90757 CATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGAT 90816
Qy 421 TGGAGGAGCTCCACTCCAGCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGATGCCC 480
Db 90817 TGGAGGAGCTCCACTCCAGCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGATGCCC 90876
Qy 481 TCTTCAGAGAGATGTTGGCTGCCACTTCATCAAGCTCGGAGCTCTCTTGGACTACTGCG 540
Db 90877 TCTTCAGAGAGATGTTGGCTGCCACTTCATCAAGCTCGGAGCTCTCTTGGACTACTGCG 90936
Qy 541 AGGCTCTAGAGAAATCCCGGGGGGAAGATTCTCGCGCTGCTCGA 584
Db 90937 AGGCTCTAGAGAAATCCCGGGGGGAAGATTCTCGCGCTGCTCGA 90980

RESULT 8
HS21C080 340000 bp DNA linear PRI 24-MAY-2000
LOCUS Homo sapiens chromosome 21 segment HS21C080.
DEFINITION AL163280 AP001735 BA000005
VERSION AL163280.2 GI:7717369
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 340000)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyma,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordliek,G.,
Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Riesemann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
Yaspo,M.L.
Direct Submission
Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing
Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
Group * Institute of Molecular Biotechnology, Genome Analysis *
Keio University School of Medicine, Dept. of Molecular Biology *
GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
Genetics (addresses see below)
The Chromosome 21 Mapping and Sequencing Consortium consists of
* RIKEN Genomic Sciences Center, Human Genome Research Group, *
Sagamihara 228-8555, Japan,
* e.mail: sakaki@gs.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
```

```
* Institute of Molecular Biotechnology, Genome Analysis, *
Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Dept. of Molecular Biology, *
Tokyo 160-8582, Japan,
* e.mail: ehimizu@mb-med.keio.ac.jp
* URL: http://adenine.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Ihnestrassse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/.
FEATURES
source
1..340000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
<1..125946
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P160L9, 5' partial"
/clone_lib="RPC11,3-5 PAC library"
/notes="Accession No. AF121897"
73410..243533
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P70124"
/clone_lib="RPC11,3-5 PAC library"
/notes="Accession No. AF064860"
224137..>340000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P206A10, 3' partial"
/clone_lib="RPC11,3-5 PAC library"
/notes="Accession No. AF121782"
286628..>340000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="BAC-291B3, 3' partial"
/clone_lib="BAC library"
/notes="Accession No. AF064857"
1088..1566
/notes="LIM4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
2398..2495
/notes="(CTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(2496..2878)
/notes="THE1C"
/rpt_family="LTR/MaLR"
```

```

repeat_region      /rpt_type=DISPERSED
2879..2931
/note="TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region      /rpt_type=TANDEM
2932..3004
/note="TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region      /rpt_type=TANDEM
3005..3074
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
gene               /rpt_type=TANDEM
3473..19157
/gene="SH3BGR"
mRNA               join(3473..3565,12649..12678,15331..15460,18739..19157)
/gene="SH3BGR"
CDS                 join(3473..3565,12649..12678,15331..15426)
/gene="SH3BGR"
/note="Accession No. X93498"
/codon_start=1
/product="21-Glutamic Acid Rich protein 21-GARP"
/protein_id="CAB90445.1"
/db_xref="GI:7717370"
/db_xref="GOA:P55822"
/translation="GSEKAEAGGTEAOKESGSDVGNLPEAQKNEEBEGTATEETIE
IAMEGAEAGEAEEEBETASGERPGEDEDS"
3473..3565
/gene="SH3BGR"
/number=4
repeat_region      5542..5829
/note="AluX"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region      complement(6887..7067)
/note="LiMD3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
repeat_region      complement(8482..8762)
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region      complement(8865..8964)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region      9131..9424
/note="AluSq"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region      complement(9723..9882)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region      complement(9989..10077)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region      complement(10078..10429)
/note="THE18"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region      complement(10430..10651)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region      10980..11000
/note="(TTTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region      complement(11002..11282)
/note="AluSp"

```

```

repeat_region      /rpt_family="SINE/Alu"
complement(11315..11556)
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
11702..11753
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
12649..12678
/gene="SH3BGR"
/number=5
repeat_region      complement(13115..13425)
/note="AluX"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region      complement(13598..13892)
Query Match      100.0%; Score 584; DB 9; Length 340000;
Best Local Similarity 100.0%; Pred. No. 7.3e-151;
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGACATTATCCAGAAGATTTCCTAGACGCTCTATTACAAATCTGACCTGAAGACCATGA 60
DB 163805 GGGACATTATCCAGAAGATTTCCTAGACGCTCTATTACAAATCTGACCTGAAGACCATGA 163864
QY 61 TGGGCATAGTAATGGTCCATCGCTTTTGTCTCAGGGGGCGTTTGTGTATGAAACAGACT 120
DB 163865 TGGGCATAGTAATGGTCCATCGCTTTTGTCTCAGGGGGCGTTTGTGTATGAAACAGACT 163924
QY 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAACAGAACAA 180
DB 163925 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAACAGAACAA 163984
QY 181 CCAGGTTTTTCACCTGGCTTCTTGAACTCAATGAGTTTCCCATCAGGAGCCATTTCAGCA 240
DB 163985 CCAGGTTTTTCACCTGGCTTCTTGAACTCAATGAGTTTCCCATCAGGAGCCATTTCAGCA 164044
QY 241 AGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGTACCCACCATTTCTGCTCCGGCA 300
DB 164045 AGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGTACCCACCATTTCTGCTCCGGCA 164104
QY 301 CCGGCTACGTGTTTCTGGGACGTCGCGAGTCAGGTGTACAATGTCTTCCAAGAGCGTCC 360
DB 164105 CCGGCTACGTGTTTCTGGGACGTCGCGAGTCAGGTGTACAATGTCTTCCAAGAGCGTCC 164164
QY 361 CATACATTAACTGGAAGACGCTGTTTGTGGGGCTCTGCTCTGAAAGGCTGAACATCAGAT 420
DB 164165 CATACATTAACTGGAAGACGCTGTTTGTGGGGCTCTGCTCTGAAAGGCTGAACATCAGAT 164224
QY 421 TGGAGGAGCTCCACTCCCGCCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC 480
DB 164225 TGGAGGAGCTCCACTCCCGCCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC 164284
QY 481 TCTTCAGGAGATCGTGGCTCTGCCCTTTCATCAAGCCCTCGGACTCTCTTGAGCTACTGGC 540
DB 164285 TCTTCAGGAGATCGTGGCTCTGCCCTTTCATCAAGCCCTCGGACTCTCTTGAGCTACTGGC 164344
QY 541 AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCGCCCTGTCTGA 584
DB 164345 AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCGCCCTGTCTGA 164388

```

```

RESULT 9
HSA6078          HSA6078          933 bp      DNA      linear      PRI 11-MAY-2000
LOCUS            Homo sapiens beta3gal-T6 gene.
DEFINITION       AJ006078
ACCESSION        AJ006078
VERSION          AJ006078.1 GI:7799922
KEYWORDS         beta-1,3-galactosyltransferase; beta3gal-T6 gene.
SOURCE           Homo sapiens (human)
ORGANISM         Homo sapiens

```



misc feature

source	1. .1576 /organism="Homo sapiens" /mol_type="genomic DNA" /isolate="#056" /db_xref="taxon:9606" /note="human sequence used for primer design based on Acc# AB020337" 1. .27 /number=3 28. .674 /number=3 675. .1576 /gene="beta1,3-Galt 5" 675. .>1576 /gene="beta1,3-Galt 5" /codon_start=1 /product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5" /protein_id="BAA94501.1" /db_xref="GI:7593027" /translation="WAPPKMLMYICLLVGLALCLYFSMYSNLNPKQSFVYKDGNF LKLPTDCRQTPPELVLLVTSKQLAERMAIRQTWGERVKGKQLKTFPLGTTSS AAETKEVDQESQRHGDIIQKDFLDGYNNLTMTMGIEWHVFPCQAAPVMTKDSMF INVDYLLELLKKNRTTRPFTGFLKNEPIRQPFKVFWSKSEYPDWRPFPFCSGTG YVFGDVASQVYNSKSPYIKLEDVFGVGLCLERLNIRLELHSHQPTFFPGGLRFSVC LFRRIVACHFIKPRTLDDYWQALE" 675. .>1576 /gene="beta1,3-Galt 5" /number=4	RESULT 13 AB041415 LOCUS DEFINITION Pan paniscus beta1,3-Galt 5 gene for UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5, partial cds. ACCESSION AB041415 VERSION AB041415.1 KEYWORDS GI:7593024 SOURCE Pan paniscus (pygmy chimpanzee) ORGANISM Pan paniscus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan. Liu, Y. and Saitou, N. 1 (bases 1 to 1570) 2 (bases 1 to 1570) Published Only in DataBase (2000) REFERENCE Liu, Y. and Saitou, N. Direct Submission Submitted (11-APR-2000) Naruya Saitou, National Institute of Genetics, Laboratory of Evolutionary Genetics; Ilii Yata, Mishima, Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp, URL:http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790, Fax:81-559-81-6789) FEATURES Location/Qualifiers 1. .1570 /organism="Pan paniscus" /mol_type="genomic DNA" /isolate="bonobo-05" /db_xref="taxon:9597" /note="human sequence used for primer design based on Acc# AB020337"
exon	1 GGGACATATCCAGAGGATTTCTTAGACGCTATTACAACTCGACCTGAAGACCATGA 60 GGGACATATCCAGAGGATTTCTTAGACGCTATTACAACTCGACCTGAAGACCATGA 1083	exon 1. .20 /number=3 21. .667 /number=3 668. .1570 /gene="beta1,3-Galt 5" 668. .>1570 /gene="beta1,3-Galt 5" /codon_start=1 /product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5" /protein_id="BAA94500.1" /db_xref="GI:7593025" /translation="WAPPKMLMYICLLVGLALCLYFSMYSNLNPKQSFVYKDGNF LKLPTDCRQTPPELVLLVTSKQLAERMAIRQTWGERVKGKQLKTFPLGTTSS AAETKEVDQESQRHGDIIQKDFLDGYNNLTMTMGIEWHVFPCQAAPVMTKDSMF INVDYLLELLKKNRTTRPFTGFLKNEPIRQPFKVFWSKSEYPDWRPFPFCSGTG YVFGDVASQVYNSKSPYIKLEDVFGVGLCLERLNIRLELHSHQPTFFPGGLRFSVC LFRRIVACHFIKPRTLDDYWQALE" 668. .>1570 /gene="beta1,3-Galt 5" /number=4
intron	124 GGGACATATCCAGAGGATTTCTTAGACGCTATTACAACTCGACCTGAAGACCATGA 1083	intron 21. .667 /number=3 668. .1570 /gene="beta1,3-Galt 5" 668. .>1570 /gene="beta1,3-Galt 5" /codon_start=1 /product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5" /protein_id="BAA94500.1" /db_xref="GI:7593025" /translation="WAPPKMLMYICLLVGLALCLYFSMYSNLNPKQSFVYKDGNF LKLPTDCRQTPPELVLLVTSKQLAERMAIRQTWGERVKGKQLKTFPLGTTSS AAETKEVDQESQRHGDIIQKDFLDGYNNLTMTMGIEWHVFPCQAAPVMTKDSMF INVDYLLELLKKNRTTRPFTGFLKNEPIRQPFKVFWSKSEYPDWRPFPFCSGTG YVFGDVASQVYNSKSPYIKLEDVFGVGLCLERLNIRLELHSHQPTFFPGGLRFSVC LFRRIVACHFIKPRTLDDYWQALE"
gene	61 TGGGCATAGATGGTCCATCGCTTTTGTCTCAGCGCGGCTTTGTGATGAACAGACT 120 1084 TGGGCATAGATGGTCCATCGCTTTTGTCTCAGCGCGGCTTTGTGATGAACAGACT 1143	gene 668. .1570 /gene="beta1,3-Galt 5" 668. .>1570 /gene="beta1,3-Galt 5" /codon_start=1 /product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5" /protein_id="BAA94500.1" /db_xref="GI:7593025" /translation="WAPPKMLMYICLLVGLALCLYFSMYSNLNPKQSFVYKDGNF LKLPTDCRQTPPELVLLVTSKQLAERMAIRQTWGERVKGKQLKTFPLGTTSS AAETKEVDQESQRHGDIIQKDFLDGYNNLTMTMGIEWHVFPCQAAPVMTKDSMF INVDYLLELLKKNRTTRPFTGFLKNEPIRQPFKVFWSKSEYPDWRPFPFCSGTG YVFGDVASQVYNSKSPYIKLEDVFGVGLCLERLNIRLELHSHQPTFFPGGLRFSVC LFRRIVACHFIKPRTLDDYWQALE"
CDS	121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTCTTCTGAAGAAAAACAGAACAA 180 1144 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTCTTCTGAAGAAAAACAGAACAA 1203 181 CCAGGTTTTTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGCGACCATTCAGCA 240 1204 CCAGGTTTTTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGCGACCATTCAGCA 1263	CDS 668. .1570 /gene="beta1,3-Galt 5" 668. .>1570 /gene="beta1,3-Galt 5" /codon_start=1 /product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5" /protein_id="BAA94500.1" /db_xref="GI:7593025" /translation="WAPPKMLMYICLLVGLALCLYFSMYSNLNPKQSFVYKDGNF LKLPTDCRQTPPELVLLVTSKQLAERMAIRQTWGERVKGKQLKTFPLGTTSS AAETKEVDQESQRHGDIIQKDFLDGYNNLTMTMGIEWHVFPCQAAPVMTKDSMF INVDYLLELLKKNRTTRPFTGFLKNEPIRQPFKVFWSKSEYPDWRPFPFCSGTG YVFGDVASQVYNSKSPYIKLEDVFGVGLCLERLNIRLELHSHQPTFFPGGLRFSVC LFRRIVACHFIKPRTLDDYWQALE"
exon	241 AGTGGTTTGTGAGTAAATCTGAATATCGTGGGACAGGTACCCACCATTCGCTCCGGCA 300 1264 AGTGGTTTGTGAGTAAATCTGAATATCGTGGGACAGGTACCCACCATTCGCTCCGGCA 1323 301 CCGGCTAGCTGTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCCTCCAGAGCGTCC 360 1324 CCGGCTAGCTGTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCCTCCAGAGCGTCC 1383	exon 668. .1570 /gene="beta1,3-Galt 5" /number=4
ORIGIN	361 CATACATTAACATGGAAGACGCTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGAT 420 1384 CATACATTAACATGGAAGACGCTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGAT 1443 421 TGGAGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC 480 1444 TGGAGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC 1503 481 TCTTCAGAGGATCGTGGCCTGCCACTTTCATCAAGCCCTCGGACTCTCTTGACTACTGGC 540 1504 TCTTCAGAGGATCGTGGCCTGCCACTTTCATCAAGCCCTCGGACTCTCTTGACTACTGGC 1563 541 AGGCTCTAGAGAA 553 1564 AGGCTCTAGAGAA 1576	ORIGIN Query Match 92.7%; Score 541.2; DB 9; Length 1570; Best Local Similarity 98.6%; Pred. No. 5.1e-139; Matches 546; Conservative 0; Mismatches 8; Indels 0; Gaps 0; QY 1 GGGACATATCCAGAGGATTTCTTAGACGCTATTACAACTCGACCTGAAGACCATGA 60 DB 1017 GGGACATATCCAGAGGATTTCTTAGACGCTATTACAACTCGACCTGAAGACCATGA 1076 QY 61 TGGGCATAGATGGTCCATCGCTTTTGTCTCAGCGGGCGTGTGATGAACAGACT 120 DB 1077 TGGGCATAGATGGTCCATCGCTTTTGTCTCAGCGGGCGTGTGATGAACAGACT 1136 QY 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTCTGCTTCTCAAGAAAAACAGAA 180 DB 1137 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTCTGCTTCTCAAGAAAAACAGAA 1196





JOURNAL

Submitted (11-APR-2000) Naruya Saitou, National Institute of  
Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima,  
Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp,  
URL: http://sayer.iab.nig.ac.jp/~silver/, Tel: 81-559-81-6790,  
Fax: 81-559-81-6789)

FEATURES

source

Location/Qualifiers  
1. .1565  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/isolate="human-NR"  
/db\_xref="taxon:9606"  
/note="human sequence used for primer design based on Acc#  
AB020337"  
<1. .27  
/number=3  
28. .674  
/number=3  
675. .1565  
/gene="beta1,3-GalT 5"  
675. .>1565  
/gene="beta1,3-GalT 5"  
/codon\_start=1  
/product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5"  
/protein\_id="BAA94498.1"  
/db\_xref="GI:7593021"  
/translation="MAPPKRLMVLICLLVGLALCLYFSWYSLNPKBQSFYVKDGNF  
LKLPTDPCQTPPELVLLVTSKQLAERMAIRQTWNGKRVKQKLPFLGLTSS  
AAETKEVDQESQRGDIIQKDFLDVYINLTUKTMGIEWHVRFCQAAPVNTKDSNMF  
INVLYLTELKLNKRTTRFFFLKLINEFFPIRQPFKSWFVSKSEYPWDRYPPFCSTG  
YVFSGDVSKSVYNYKLEDFVFLGCLERLNIRLELHSQPTFFPFGGLRFSVC  
LFRIVACHFIKPRTLMTGR"  
675. .>1565  
/gene="beta1,3-GalT 5"  
/number=4

exon

intron

gene

CDS

Db 1504 TCCTCAGGAGGATCGTGGCTCGCACTTCATCAAGCCTCGACTC-TCGACTACTGGC 1562  
Qy 541 AGG 543  
Db 1563 AGG 1565

Search completed: September 22, 2005, 14:50:07  
Job time : 2673.69 secs

ORIGIN

Query Match 90.9%; Score 531; DB 9; Length 1565;  
Best Local Similarity 99.8%; Pred. No. 3.4e-136;  
Matches 542; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
Qy 1 GGGACATTATCCAGAAGGATTTCCTAGACGTCATTACAATCTGACCCCTGAAGACCATGA 60  
Db 1024 GGGACATTATCCAGAAGGATTTCCTAGACGTCATTACAATCTGACCCCTGAAGACCATGA 1083  
Qy 61 TGGGCATAGATGGTCCATCGCTTTTGTCTCAGCGCGCGTCTTGTGATGAACACAGACT 120  
Db 1084 TGGGCATAGATGGTCCATCGCTTTTGTCTCAGCGCGCGTCTTGTGATGAACACAGACT 1143  
Qy 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAA 180  
Db 1144 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAA 1203  
Qy 181 CCAGGTTTTTCATGCGCTTCTGAACTCAATAGTATTTCCATCAGGCGCCATTCAGCA 240  
Db 1104 CCAGGTTTTTCATGCGCTTCTGAACTCAATAGTATTTCCATCAGGCGCCATTCAGCA 1263  
Qy 241 AGTGGTTTGTGAGTAAATCTGAATATCGTGGGACAGGTACCCACCATTTCTGCTCCGGCA 300  
Db 1164 AGTGGTTTGTGAGTAAATCTGAATATCGTGGGACAGGTACCCACCATTTCTGCTCCGGCA 1323  
Qy 301 CCGGCTACGTGTTTCTGGCGACGTGGCGAGTCAAGTGTACAATGTCTCCAAAGAGCGTCC 360  
Db 1124 CCGGCTACGTGTTTCTGGCGACGTGGCGAGTCAAGTGTACAATGTCTCCAAAGAGCGTCC 1383  
Qy 361 CATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGAT 420  
Db 1184 CATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGAT 1443  
Qy 421 TGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTAGCGTTCTCCGTTATGCC 480  
Db 1444 TGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTAGCGTTCTCCGTTATGCC 1503  
Qy 481 TCCTCAGGAGGATCGTGGCTCGCCACTTTCATCAAGCCTCGGACTCTCTTGGACTACTGGC 540

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2005, 05:52:47 ; Search time 348.351 Seconds  
(without alignments)  
9924.272 Million cell updates/sec

Title: US-10-777-828-8\_COPY\_428\_1011

Perfect score: 584

Sequence: 1 gggacattatccagaagat.....aagattgccgcgtgtctga 584

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004as:\*  
13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	584	100.0	933	12 ADK68456	Adk68456 Human bet
2	584	100.0	1011	3 AA27959	Aa27959 Human bet
3	584	100.0	2629	10 ABZ56918	Abz56918 B3GALT nu
4	584	100.0	2762	10 ABZ56917	Abz56917 B3GALT nu
5	584	100.0	2775	3 AA93875	Aa93875 Human bet
6	584	100.0	3409	10 ABZ56919	Abz56919 B3GALT nu
7	584	100.0	10562	3 AA93876	Aa93876 Human bet
8	583	99.8	933	10 ABZ56916	Abz56916 B3GALT nu
9	114.4	19.6	1266	2 AA35710	Aa35710 cDNA enco
10	114.4	19.6	1268	10 ABZ56911	Abz56911 B3GALT nu
11	114.4	19.6	1269	8 AD48744	Aa48744 Murine be
12	114.4	19.6	1269	10 ABZ56913	Abz56913 B3GALT nu
13	114.4	19.6	1473	12 ADM66890	Adm66890 Murine ad
14	114.4	19.6	2420	2 AA35711	Aa35711 cDNA enco
15	114.4	19.6	3054	12 ADM66891	Adm66891 Human adi
16	114.4	19.6	3212	10 ABZ56914	Abz56914 B3GALT nu
17	114.4	19.6	3215	10 ABZ56912	Abz56912 B3GALT nu
18	114.4	19.6	3215	10 ABZ56915	Abz56915 B3GALT nu
19	110.6	18.9	1037	10 ABZ56909	Abz56909 B3GALT nu
20	110.6	18.9	1739	2 AAQ67067	Aa67067 Beta-1,3-

## ALIGNMENTS

## RESULT 1

ID ADK68456 standard; cDNA; 933 BP.

AC ADK68456;

DT 06-MAY-2004 (first entry)

DE Human beta-1,3-galactosyl transferase cDNA SeqID 1.

KW human; gene; ss; saccharide binding protein; maltose binding protein;  
KW beta-1,3-galactosyl transferase; inflammation; infectious disease;  
KW cancer metastasis suppression; dairy product; antiinflammatory;  
KW antimicrobial; cytostatic.

OS Homo sapiens.

XX Key Location/Qualifiers

FH CDS 1..933

FT /\*tag= a

FT /product= "Beta-1,3-galactosyl transferase protein"

XX JP2004016117-A.

XX 22-JAN-2004.

XX 17-JUN-2002; 2002JP-00176132.

XX 17-JUN-2002; 2002JP-00176132.

XX (TOYM ) TOYOBO KK.

XX WPI; 2004-113878/12.

XX P-PSDB; ADK68457.

XX New fusion protein useful in diagnosis of diseases and in manufacture of pharmaceutical products, comprises a saccharide binding protein and beta 1, 3-galactosyl transferase.

XX Example 1; SEQ ID NO 1; 23pp; Japanese.

XX This invention relates to a novel recombinant fusion protein that comprises a saccharide binding protein, in particular a maltose binding

21	109	18.7	2168	10	ABZ56910	Abz56910 B3GALT nu
22	106.2	18.2	1092	12	AD000390	Ado00390 Novel hum
23	106.2	18.2	1092	12	ADN98821	Adn98821 Novel hum
24	106.2	18.2	1358	10	ADF76806	Adf76806 Novel hum
25	106.2	18.2	1773	2	AA87193	Aa87193 Human Den
26	106.2	18.2	1897	6	ABK51201	Abk51201 Human cDN
27	106.2	18.2	1897	12	ADO19801	Ado19801 Human PRO
28	106.2	18.2	2024	9	ADA20069	Ada20069 Novel hum
29	106.2	18.2	2095	3	AAZ65022	Aaz65022 Membrane-
30	106.2	18.2	2095	4	AA846005	Aa846005 Human DNA
31	106.2	18.2	2095	4	AA92075	Aa92075 Human PRO
32	106.2	18.2	2095	5	AA44168	Aa44168 Human PRO
33	106.2	18.2	2095	6	AB574395	Ab574395 Human cDN
34	106.2	18.2	2095	8	ACA89455	Aca89455 cDNA enco
35	106.2	18.2	2095	8	ACA73465	Aca73465 Human sec
36	106.2	18.2	2095	8	ACA05780	Aca05780 Human sec
37	106.2	18.2	2095	8	ACA66614	Aca66614 cDNA enco
38	106.2	18.2	2095	8	ACA64304	Aca64304 Novel hum
39	106.2	18.2	2095	8	ACA91181	Aca91181 Novel hum
40	106.2	18.2	2095	8	ACD81558	Acd81558 Human cDN
41	106.2	18.2	2095	8	ACF20189	Acf20189 Human sec
42	106.2	18.2	2095	8	ACF19575	Acf19575 Human sec
43	106.2	18.2	2095	8	ACD21863	Acd21863 Human sec
44	106.2	18.2	2095	8	ACF13028	Acf13028 Human sec
45	106.2	18.2	2095	8	ACD25131	Acd25131 Human sec

CC protein, and a beta-1,3-galactosyl transferase and an appropriate  
 CC manufacturing method. Specifically, it refers to fusion protein that can  
 CC transfer galactose to an N-acetyl glucosamine residue or N-acetyl  
 CC glucosamine monosaccharide. The present invention describes a method to  
 CC produce this enzyme, cheaply and efficiently, for the diagnosis and  
 CC treatment of inflammation, infectious diseases or for cancer metastasis  
 CC suppression. Furthermore, it can also be useful for improving dairy  
 CC products. Accordingly, compositions exhibit various activities including  
 CC anti-inflammatory, antimicrobial and cytostatic. This polynucleotide  
 CC sequence is the human beta-1,3-galactosyl transferase cDNA of the  
 CC invention.

XX SQ Sequence 933 BP; 227 A; 232 C; 241 G; 233 T; 0 U; 0 Other;

Query Match 100.0%; Score 584; DB 12; Length 933;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-181;  
 Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGACATTATCCAGAAAGATTTCCTAGACGCTATTACAACTGACCGTGAAGACCATGA 60  
 DB 350 GGGACATTATCCAGAAAGATTTCCTAGACGCTATTACAACTGACCGTGAAGACCATGA 409  
 QY 61 TGGGCATAGAAATGGGTCCATCGCTTTTCTCAGCGCGCGTTTGTGATGAAACAGACT 120  
 DB 410 TGGGCATAGAAATGGGTCCATCGCTTTTCTCAGCGCGCGTTTGTGATGAAACAGACT 469  
 QY 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACCTGCTCTGAAGAAACAGAACAA 180  
 DB 470 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACCTGCTCTGAAGAAACAGAACAA 529  
 QY 181 CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCCATCAGGCGAGCCATTTCAGCA 240  
 DB 530 CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCCATCAGGCGAGCCATTTCAGCA 589  
 QY 241 AGTGTGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTACCCACCAATTCGTCTCCGCA 300  
 DB 590 AGTGTGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTACCCACCAATTCGTCTCCGCA 649  
 QY 601 CCGGTACGTGTTTCTGGGACGTTGGCGAGTGGCGAGTGTACAAATGCTCCTCAAGAGCGTCC 360  
 DB 650 CCGGTACGTGTTTCTGGGACGTTGGCGAGTGGCGAGTGTACAAATGCTCCTCAAGAGCGTCC 709  
 QY 661 CATACATTAAATCGAAGACGCTGTTTGTGGGCTCTGCTCTGAAAGGCTGAACATCAGAT 420  
 DB 710 CATACATTAAATCGAAGACGCTGTTTGTGGGCTCTGCTCTGAAAGGCTGAACATCAGAT 769  
 QY 421 TGGAGGAGCTCCACTCCAGCCGACCTTTTTTCCAGGGGGCTTACGGCTTCTCCGTATGCC 480  
 DB 470 TGGAGGAGCTCCACTCCAGCCGACCTTTTTTCCAGGGGGCTTACGGCTTCTCCGTATGCC 829  
 QY 481 TCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGC 540  
 DB 830 TCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGC 889  
 QY 541 AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTGCGCCGTGCTGA 584  
 DB 890 AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTGCGCCGTGCTGA 933

RESULT 2  
 ID AAA27959 standard; DNA; 1011 BP.  
 XX AAA27959  
 AC AAA27959  
 XX  
 DT 15-AUG-2000 (first entry)  
 XX Human beta3Gal-T5 gene sequence.  
 XX UDP-D-galactose; beta-N-acetylglucosamine beta1,3-galactosyltransferase;  
 KW Beta3Gal-T5; ss; human; chromosome 21q22.3; galactosylation;  
 KW beta1,3-galactosyl glycosylated saccharide production; glycopeptide;  
 KW glycoprotein.

XX Homo sapiens.  
 OS  
 XX Key Location/Qualifiers  
 FH CDS 79..1011  
 FT /\*tag= a  
 FT /product= "Beta3Gal-T5"  
 FT complement(79..98)  
 FT primer\_bind /\*tag= b  
 FT primer\_bind complement(150..170)  
 FT /\*tag= c  
 FT primer\_bind 991..1011  
 FT /\*tag= d  
 XX  
 XX WO200029558-A1.  
 EN 25-MAY-2000.  
 XX  
 XX 11-NOV-1999; 99WO-US026807.  
 XX  
 XX 13-NOV-1998; 98DK-00001483.  
 XX  
 XX (CLAU/) CLAUSEN H.  
 XX  
 XX Clausen H, Amado M;  
 XX  
 XX WPI; 2000-399728/34.  
 DR P-PSDB; AAY94641.  
 XX  
 XX Novel nucleic acid sequence encoding human UDP-galactose:beta-N-  
 FT acetylglucosamine beta1,3-galactosyltransferase useful for obtaining beta  
 FT 1,3-galactosyl glycosylated saccharides and glycopeptides or  
 FT glycoproteins.  
 XX  
 XX Claim 7; Fig 1; 74pp; English.  
 XX  
 XX The present invention relates to a nucleic acid sequence encoding UDP-D-  
 CC galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase  
 CC (beta3Gal-T5). Beta3 transferases add galactose to the hydroxy group at  
 CC carbon 3 of 2-acetamido-2-deoxy-D-glucose (GlcNAc). The present sequence  
 CC represents the human beta3Gal-T5 gene sequence. The beta3Gal-T5 gene is  
 CC located on human chromosome 21q22.3. Beta3Gal-T5 is a type II  
 CC transmembrane glycoprotein. The invention also relates to the beta3Gal-T5  
 CC protein sequence, a nucleic acid vector comprising the beta3Gal-T5  
 CC nucleotide sequence, a host cell comprising the vector, and a method for  
 CC the production of the beta3Gal-T5 protein from the host cells. The  
 CC methods of the invention can be used for recombinant production of  
 CC beta3Gal-T5 for use as a catalyst and for recombinant production of  
 CC peptides or proteins with appropriate galactosylation. The beta3Gal-T5  
 CC protein can be used to obtain beta1,3-galactosyl glycosylated  
 CC saccharides, glycopeptides or glycoproteins

XX SQ Sequence 1011 BP; 247 A; 256 C; 251 G; 257 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 584; DB 3; Length 1011;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-181;  
 Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGACATTATCCAGAAAGATTTCCTAGACGCTCTATTACAATCTGACCCCTGAAGACCATGA 60  
 DB 428 GGGACATTATCCAGAAAGATTTCCTAGACGCTCTATTACAATCTGACCCCTGAAGACCATGA 487  
 QY 61 TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGCGCGGCTTGTGATGAAACAGACT 120  
 DB 488 TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGCGCGGCTTGTGATGAAACAGACT 547  
 QY 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAACAGAACAA 180  
 DB 548 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAACAGAACAA 607  
 QY 181 CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCCATCAGGCGAGCCATTTCAGCA 240  
 DB 608 CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCCATCAGGCGAGCCATTTCAGCA 667

QY 241 ACTGTTTCTCAGTAAATCTGAATATCCGTGGACAGGTACCCACCAATTCCTCTCCGGCA 300  
 DB 668 AGTGGTTTCTCAGTAAATCTGAATATCCGTGGACAGGTACCCACCAATTCCTCTCCGGCA 727  
 QY 301 CCGGCTACGTGTTTCTGGCAGCGTGGCAGTCAGGTGTACAAATGTCTCCAGAGCGTCC 360  
 DB 728 CCGGCTACGTGTTTCTGGCAGCGTGGCAGTCAGGTGTACAAATGTCTCCAGAGCGTCC 787  
 QY 361 CATACATTAATCTGGAAGACGTGTTTGTGGGCTCTGCTCCGAAAGGCTGAACATCAGAT 420  
 DB 788 CATACATTAATCTGGAAGACGTGTTTGTGGGCTCTGCTCCGAAAGGCTGAACATCAGAT 847  
 QY 421 TGGAGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC 480  
 DB 848 TGGAGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC 907  
 QY 481 TCTTCAGGAGGATCGTGGCTCCCACTTCATCAAGCTCGGACTCTCTTGGACTACTGCC 540  
 DB 908 TCTTCAGGAGGATCGTGGCTCCCACTTCATCAAGCTCGGACTCTCTTGGACTACTGCC 967  
 QY 541 AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCCGCTGTCTGA 584  
 DB 968 AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCCGCTGTCTGA 1011

RESULT 3  
 ID ABZ56918 standard; DNA; 2629 BP.  
 XX AC ABZ56918;  
 XX DT 04-APR-2003 (first entry)  
 XX DE B3GALT nucleic acid sequence # SEQ ID 10.  
 XX KW B3GALT; beta-1,3 galactosyltransferase; p53; cancer; breast; colon;  
 XX KW kidney; lung; ovary; gene; ds.  
 XX OS Homo sapiens.  
 XX FN WO200299044-A2.  
 XX PD 12-DEC-2002.  
 XX PF 02-JUN-2002; 2002WO-US017356.  
 XX PR 05-JUN-2001; 2001US-0296076P.  
 XX PR 10-OCT-2001; 2001US-0328605P.  
 XX PR 15-FEB-2002; 2002US-0357253P.  
 XX PA (EXEL-) EXELIXIS INC.  
 XX PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
 XX WPI; 2003-156849/15.

PT Identifying p53 pathway modulating agents with B3GALT genes, useful for  
 PT the diagnosis and treatment of disorders associated with defects in the  
 PT p53 pathway, such as cancer of the breast, colon, kidneys, lung and  
 PT ovary.  
 XX Disclosure; Page 53-55; 82pp; English.  
 XX CC The invention relates to identifying a candidate p53 pathway modulating  
 CC agent in humans that is referred to in the specification as BGALT (beta-  
 CC 1,3 galactosyltransferase). Also disclosed is a method for diagnosing a  
 CC disease in a patient, by contacting a sample with a probe for B3GALT  
 CC expression, and comparing the results with a control, and determining  
 CC whether the results indicate a likelihood of disease. Methods and  
 CC compositions of the invention are useful for the diagnosis and treatment  
 CC of disorders associated with defects in the p53 pathway, such as cancer  
 CC of the breast, colon, kidneys, lung and ovary. The current sequence

CC represents a B3GALT nucleic acid sequence referred to in the disclosure  
 CC of the invention  
 XX SQ Sequence 2629 BP; 632 A; 665 C; 644 G; 688 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 584; DB 10; Length 2629;  
 Best Local Similarity 100.0%; Pred. No. 9e-181;  
 Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGGACATTTATCCAGAAGGATTTCTTAGACGCTATTACAACTGACCCCTGAAGACCATCA 60  
 DB 650 GGGACATTTATCCAGAAGGATTTCTTAGACGCTATTACAACTGACCCCTGAAGACCATCA 709  
 QY 61 TGGGCATAGAATGGGTCCATCGCTTTTGTCTCAGCGCGCGTTCCTGATGAAGACCA 120  
 DB 710 TGGGCATAGAATGGGTCCATCGCTTTTGTCTCAGCGCGCGTTCCTGATGAAGACCA 769  
 QY 121 CAGACATGTTTATCAATGTTGACTATCTGACTGAATCTGTTCTGAAGAAAAACAGAA 180  
 DB 770 CAGACATGTTTATCAATGTTGACTATCTGACTGAATCTGTTCTGAAGAAAAACAGAA 829  
 QY 181 CCAGGTTTTTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGCGAGCCATTCAGCA 240  
 DB 830 CCAGGTTTTTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGCGAGCCATTCAGCA 889  
 QY 241 AGTGGTTTGTCACTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCCTCTCCGCA 300  
 DB 890 AGTGGTTTGTCACTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCCTCTCCGCA 949  
 QY 301 CCGGCTACGTGTTTCTGGCGACGTGGCGAGTCAAGTGTACAATGTCTCCAAAGAGCGTCC 360  
 DB 950 CCGGCTACGTGTTTCTGGCGACGTGGCGAGTCAAGTGTACAATGTCTCCAAAGAGCGTCC 1009  
 QY 361 CATACATTAATCTGGAAGACGTGTTTGTGGGCTCTGCTCGAAAGGCTGAACATCAGAT 420  
 DB 1010 CATACATTAATCTGGAAGACGTGTTTGTGGGCTCTGCTCGAAAGGCTGAACATCAGAT 1069  
 QY 421 TGGAGGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGGCTTACGCTTCTCCGTATGCC 480  
 DB 1070 TGGAGGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGGCTTACGCTTCTCCGTATGCC 1129  
 QY 481 TCTTCAGGAGGATCGTGGCTCGCCACTTCATCAAGCCCTCGGACTCTCTTGGACTACTGCC 540  
 DB 1130 TCTTCAGGAGGATCGTGGCTCGCCACTTCATCAAGCCCTCGGACTCTCTTGGACTACTGCC 1189  
 QY 541 AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCCGCTGTCTGA 584  
 DB 1190 AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCCGCTGTCTGA 1233

RESULT 4  
 ABZ56917  
 ID ABZ56917 standard; DNA; 2762 BP.  
 XX AC ABZ56917;  
 XX DT 04-APR-2003 (first entry)  
 XX DE B3GALT nucleic acid sequence # SEQ ID 9.  
 XX KW B3GALT; beta-1,3 galactosyltransferase; p53; cancer; breast; colon;  
 XX KW kidney; lung; ovary; gene; ds.  
 XX OS Homo sapiens.  
 XX FN WO200299044-A2.  
 XX PD 12-DEC-2002.  
 XX PF 02-JUN-2002; 2002WO-US017356.  
 XX PR 05-JUN-2001; 2001US-0296076P.  
 XX PR 10-OCT-2001; 2001US-0328605P.

```
PR 15-FEB-2002; 2002US-0357253P.
XX (EXEL-) EXELIXIS INC.
XX
XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
XX WPI; 2003-156849/15.
XX
XX Identifying p53 pathway modulating agents with B3GALT genes, useful for
XX the diagnosis and treatment of disorders associated with defects in the
XX p53 pathway, such as cancer of the breast, colon, kidneys, lung and
XX ovary.
XX
XX Example 5; Page 52-53; 82pp; English.
XX
XX The invention relates to identifying a candidate p53 pathway modulating
XX agent in humans that is referred to in the specification as BGALT (beta-
XX 1,3 galactosyltransferase). Also disclosed is a method for diagnosing a
XX disease in a patient, by contacting a sample with a probe for B3GALT
XX expression, and comparing the results with a control, and determining
XX whether the results indicate a likelihood of disease. Methods and
XX compositions of the invention are useful for the diagnosis and treatment
XX of disorders associated with defects in the p53 pathway, such as cancer
XX of the breast, colon, kidneys, lung and ovary. The current sequence
XX represents a B3GALT nucleic acid sequence referred to in an example from
XX the invention
XX
XX Sequence 2762 BP; 651 A; 703 C; 679 G; 729 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 584; DB 10; Length 2762;
XX Best Local Similarity 100.0%; Pred. No. 9.2e-181;
XX Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGACATTATCCAGAAGGATTTCCTAGACGCTCTATTACAATCTGACCTGAAGACCATGA 60
DB 783 GGGACATTATCCAGAAGGATTTCCTAGACGCTCTATTACAATCTGACCTGAAGACCATGA 842
QY 61 TGGGCATAGATGGTCCATCGCTTTCTCTCAGCGCGCTTTGTGATGAACACAGACT 120
DB 843 TGGGCATAGATGGTCCATCGCTTTCTCTCAGCGCGCTTTGTGATGAACACAGACT 902
QY 121 CAGACATCTTCATCAATCTTGCACTATCTGACTCTGAACTCTTCTGAAGAAAACAGAACAA 180
DB 903 CAGACATCTTCATCAATCTTGCACTATCTGACTGAACTCTTCTGAAGAAAACAGAACAA 962
QY 181 CCAGGTTTTTCATCGGCTTCTTGAACTCAATAGATTTTCCATCAGGAGCCATTCAGCA 240
DB 963 CCAGGTTTTTCATCGGCTTCTTGAACTCAATAGATTTTCCATCAGGAGCCATTCAGCA 1022
QY 241 AGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTACCCACATTCGCTCCGGCA 300
DB 1023 AGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTACCCACATTCGCTCCGGCA 1082
QY 301 CCGGCTAGCTGTTTCTGGCGACGTGGCGAGTCTAGGTGTACAATGTCTCCAGAGCGTCC 360
DB 1083 CCGGCTAGCTGTTTCTGGCGACGTGGCGAGTCTAGGTGTACAATGTCTCCAGAGCGTCC 1142
QY 361 CATACATTAAATCGAAGACGTTGTTGTGGGGCTCTGCCTCGAAAGGTGCAACATCAGAT 420
DB 1143 CATACATTAAATCGAAGACGTTGTTGTGGGGCTCTGCCTCGAAAGGTGCAACATCAGAT 1202
QY 421 TGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTCTCCGATGCC 480
DB 1203 TGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTCTCCGATGCC 1262
QY 481 TCTTCAGGAGGATCGTGGCGCTGCCACTTTCATCAAGCTCGGACTCTCTTGACTACTGGC 540
DB 1263 TCTTCAGGAGGATCGTGGCGCTGCCACTTTCATCAAGCTCGGACTCTCTTGACTACTGGC 1322
QY 541 AGGCTCTAGAGAAATCCCGGGGGGGAAGATGTTCGCCCTGTCTGA 584
DB 1323 AGGCTCTAGAGAAATCCCGGGGGGGAAGATGTTCGCCCTGTCTGA 1366
```

```
RESULT 5
AAA93875
ID AAA93875 standard; DNA; 2775 BP.
XX
XX AAA93875;
XX
XX 15-JAN-2001 (first entry)
XX
XX Human beta-1,3 galactose transferase encoding DNA.
XX
XX Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
XX digestive system; ds.
XX
XX Homo sapiens.
XX
XX WO2000050608-A1.
XX
XX 31-AUG-2000.
XX
XX 24-FEB-2000; 2000WO-JP001070.
XX
XX 25-FEB-1999; 99JP-00047571.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX
XX Narimatsu H, Ieshiki S, Togayachi A, Sasaki K;
XX
XX WPI; 2000-549409/50.
XX
XX P-PSDB; AAB93875.
XX
XX Beta-1,3 galactose transferase and DNA encoding it, useful for synthesis
XX of type 1 sialyl Lewis, a carbohydrate for treatment of digestive system
XX cancer.
XX
XX Claim 5; Page 99-102; 123pp; Japanese.
XX
XX This invention relates to a polypeptide (I) with beta-1,3 galactose
XX transferase activity, or variants of (I) comprising amino acid additions,
XX deletions and/or substitutions. Included in the invention is DNA encoding
XX all or part of (I); expression vectors containing the DNA, host cells
XX transformed by the vectors; a method for the preparation of the
XX polypeptide by culture of the transformants or by expression of the
XX of a transgenic mammal, and antibodies recognising (I). The Beta-1,3
XX galactose transferase protein transfers galactose by beta-1,3 bonding to
XX N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as
XX GlcNAc) to give Galbeta1-4Glc to give Galbeta1-3GlcNAc. The protein and DNA
XX encoding it are useful for the treatment and diagnosis of cancer of the
XX digestive system. The present sequence represents Beta-1,3 galactose
XX transferase encoding DNA
XX
XX Sequence 2775 BP; 681 A; 698 C; 669 G; 727 T; 0 U; 0 Other;
```

```
Query Match 100.0%; Score 584; DB 3; Length 2775;
Best Local Similarity 100.0%; Pred. No. 9.2e-181;
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGACATTATCCAGAAGATTTCCTAGACGCTCTATTACAATCTGACCTGAAGACCATGA 60
DB 751 GGGACATTATCCAGAAGATTTCCTAGACGCTCTATTACAATCTGACCTGAAGACCATGA 810
QY 61 TGGGCATAGATGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAACACAGACT 120
DB 811 TGGGCATAGATGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAACACAGACT 870
QY 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAAACAGAACAA 180
DB 871 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAAACAGAACAA 930
QY 181 CCAGGTTTTTCATCGGCTTCTTGAACTCAATGAGTTTCCATCAGGAGCCATTCAGCA 240
DB 931 CCAGGTTTTTCATCGGCTTCTTGAACTCAATGAGTTTCCATCAGGAGCCATTCAGCA 990
```

CC	of the invention
XX	
SQ	Sequence 3409 BP; 833 A; 842 C; 833 G; 901 T; 0 U; 0 Other;
	Query Match      100.0%; Score 584; DB 10; Length 3409;
	Best Local Similarity 100.0%; Pred. No. 1e-180;
	Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 GGACATTCAGAGGATTCTTAGAGGTCTATTACAATCTGACCCTGAACCATGA 60 
Db	1430 GGACATTCAGAGGATTCTTAGAGGTCTATTACAATCTGACCCTGAACCATGA 1489 
Qy	61 TGCGCATAGAATGGGTC CATCGCTTTTGTCTCAGCGCGGCTTTTGATGAAACAGACT 120 
Db	1490 TGCGCATAGAATGGGTC CATCGCTTTTGTCTCAGCGCGGCTTTTGATGAAACAGACT 1549 
Qy	121 CGACATGTTTCATCAATGTTGACTATCTGACTGAAC TGCTCTCGRAGAAAACAGAACAA 180 
Db	1550 CGACATGTTTCATCAATGTTGACTATCTGACTGAAC TGCTCTCGRAGAAAACAGAACAA 1609 
Qy	181 CCAGGTTTTTCATCGGCTTCTTGAAACTCAATGAGT TTCCCATCAGCGACGCATTCAGCA 240 
Db	1610 CCAGGTTTTTCATCGGCTTCTTGAAACTCAATGAGT TTCCCATCAGCGACGCATTCAGCA 1669 

Qy	241	AGTGGTTGTGACAGTAAATCTGAAATATCCGTGGGACAGGTACCCACCATTCTGCTCGGGCA	1729
Db	1670	AGTGGTTGTGACAGTAAATCTGAAATATCCGTGGGACAGGTACCCACCATTCTGCTCGGGCA	1729
Qy	301	CCGGCTACGTGTTTCTGGCGACGTCGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCC	360
Db	1730	CCGGCTACGTGTTTCTGGCGACGTCGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCC	1789
Qy	361	CATACATTAACTGGAAGACGTGTTTGTGGGGCTTGCCTCGAAAGCGTGAACATCAGAT	420
Db	1790	CATACATTAACTGGAAGACGTGTTTGTGGGGCTTGCCTCGAAAGCGTGAACATCAGAT	1849
Qy	421	TGGAGGAGCTCCACTCCCGACCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC	480
Db	1850	TGGAGGAGCTCCACTCCCGACCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC	1909
Qy	481	TCCTTCAGGAGGATCGTGGGCTTGCACCTTCATCAAGCGCTCGGACTCTCTTGGACTACTGGC	540
Db	1910	TCCTTCAGGAGGATCGTGGGCTTGCACCTTCATCAAGCGCTCGGACTCTCTTGGACTACTGGC	1969
Qy	541	AGGCTCTAGAGAATTCCCGGGGGGAAGATTGTCCGCGCTGTCTGA	584
Db	1970	AGGCTCTAGAGAATTCCCGGGGGGAAGATTGTCCGCGCTGTCTGA	2013

RESULT 7  
 AAA93876  
 ID AAA93876 standard; DNA; 10562 BP.  
 XX  
 AC AAA93876;  
 XX  
 DT 15-JAN-2001 (first entry)  
 XX  
 XX Human beta3Gal-T5 encoding DNA.  
 XX  
 KW Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;  
 KW digestive system; beta3Gal-T5; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO20050608-A1.  
 XX  
 PD 31-AUG-2000.  
 XX  
 PD 24-FEB-2000; 2000WO-JP001070.  
 XX  
 PR 25-FEB-1999; 99JP-00047571.  
 XX  
 PA (KYOW ) KYOWA HAKKO KOGYO KK.

XX Narinatsu H, Ieshiki S, Togayachi A, Sasaki K;  
 PI WPI; 2000-549409/50.  
 DR  
 XX Beta-1,3 galactose transferase and DNA encoding it, useful for synthesis  
 PT of type 1 sialyl Lewis, a carbohydrate for treatment of digestive system  
 PT cancer.  
 XX  
 PS Claim 31; Page 103-111; 123pp; Japanese.  
 XX  
 CC This invention relates to a polypeptide (I) with beta-1,3 galactose  
 CC transferase activity, or variants of (I) comprising amino acid additions,  
 CC deletions and/or substitutions. Included in the invention is DNA encoding  
 CC all or part of (I); expression vectors containing the DNA, host cells  
 CC transformed by the vectors; a method for the preparation of the  
 CC polypeptide by culture of the transformants or by expression in the milk  
 CC of a transgenic mammal, and antibodies recognising (I). The Beta-1,3  
 CC galactose transferase protein transfers galactose by beta-1,3 bonding to  
 CC N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as  
 CC GlcNAc beta1-3galactai-4glc) to give Galbeta1-3GlcNAc. The protein and DNA  
 CC encoding it are useful for the treatment and diagnosis of cancer of the  
 CC digestive system. The present sequence represents a Beta3Gal-T5 encoding  
 CC DNA sequence  
 XX  
 SQ Sequence 10562 BP; 2610 A; 2415 C; 2574 G; 2963 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 584; DB 3; Length 10562;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-180;  
 Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGGACATTATCCAGAAGATTCTCTAGACGCTATTACAACTCAGCCCTGAAGACCATGA 60  
 DB 8583 GGGACATTATCCAGAAGATTCTCTAGACGCTATTACAACTCAGCCCTGAAGACCATGA 8642  
 QY 61 TGGGCATAGATGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAACAGACT 120  
 DB 8643 TGGGCATAGATGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAACAGACT 8702  
 QY 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGCTTTCTGAAGAAAACAGAACAA 180  
 DB 8703 CAGACATGTTTCATCAATGTTGACTATCTGACTGCTTTCTGAAGAAAACAGAACAA 8762  
 QY 181 CCAGGTTTTTCACTGGCTCTTGAACCTCATGATGTTCCCATCAGGCGCCATTCAGCA 240  
 DB 8763 CCAGGTTTTTCACTGGCTCTTGAACCTCATGATGTTCCCATCAGGCGCCATTCAGCA 8822  
 QY 241 AGTGGTTTGTGATGAAATCTGAATATCCGTGGGACAGGTACCCACCATTCGTCTCCGGCA 300  
 DB 8823 AGTGGTTTGTGATGAAATCTGAATATCCGTGGGACAGGTACCCACCATTCGTCTCCGGCA 8882  
 QY 301 CCGGCTACGTGTTTCTGGGACGTGGCGAGTCAAGTGTACAATGTTCTCCAAAGAGCGTCC 360  
 DB 8883 CCGGCTACGTGTTTCTGGGACGTGGCGAGTCAAGTGTACAATGTTCTCCAAAGAGCGTCC 8942  
 QY 361 CATACATTAACTGGAAGACGTGTTTGGGGCTCTGCCTCGAAGGCTGAACATCAGAT 420  
 DB 8943 CATACATTAACTGGAAGACGTGTTTGGGGCTCTGCCTCGAAGGCTGAACATCAGAT 9002  
 QY 421 TGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTCTCCGATATGCC 480  
 DB 9003 TGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTCTCCGATATGCC 9062  
 QY 481 TCTTCAGAGGAGTCTGGCCCTGCCACTTCATCAAGCCTCGGACTCTCTTGAAGTACTGGC 540  
 DB 9063 TCTTCAGAGGAGTCTGGCCCTGCCACTTCATCAAGCCTCGGACTCTCTTGAAGTACTGGC 9122  
 QY 541 AGGCTCTAGAGATTCCCGGGGGGAGATTGTCGCGCTGTCTGA 584  
 DB 9123 AGGCTCTAGAGATTCCCGGGGGGAGATTGTCGCGCTGTCTGA 9166

ABZ56916  
 ID ABZ56916 standard; DNA; 933 BP.  
 AC ABZ56916;  
 XX  
 DT 04-APR-2003 (first entry)  
 XX  
 DE B3GALT nucleic acid sequence # SEQ ID 8.  
 XX  
 KW B3GALT; beta-1,3 galactosyltransferase; p53; cancer; breast; colon;  
 KW kidney; lung; ovary; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200299044-A2.  
 XX  
 PD 12-DEC-2002.  
 XX  
 XX 02-JUN-2002; 2002WO-US017356.  
 PF  
 XX 05-JUN-2001; 2001US-0296076P.  
 PR  
 PR 10-OCT-2001; 2001US-0328605P.  
 PR  
 PR 15-FEB-2002; 2002US-0357253P.  
 XX  
 XX (EXEL-) EXELIXIS INC.  
 PA  
 XX  
 XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
 PI  
 XX WPI; 2003-156849/15.  
 DR  
 XX  
 XX  
 PT Identifying p53 pathway modulating agents with B3GALT genes, useful for  
 PT the diagnosis and treatment of disorders associated with defects in the  
 PT p53 pathway, such as cancer of the breast, colon, kidneys, lung and  
 PT ovary.  
 XX  
 XX  
 PS Disclosure; Page 51-52; 82pp; English.  
 XX  
 CC The invention relates to identifying a candidate p53 pathway modulating  
 CC agent in humans that is referred to in the specification as BGALT (beta-  
 CC 1,3 galactosyltransferase). Also disclosed is a method for diagnosing a  
 CC disease in a patient, by contacting a sample with a probe for B3GALT  
 CC expression, and comparing the results with a control, and determining  
 CC whether the results indicate a likelihood of disease. Methods and  
 CC compositions of the invention are useful for the diagnosis and treatment  
 CC of disorders associated with defects in the p53 pathway, such as cancer  
 CC of the breast, colon, kidneys, lung and ovary. The current sequence  
 CC represents a B3GALT nucleic acid sequence referred to in the disclosure  
 CC of the invention  
 XX  
 SQ Sequence 933 BP; 229 A; 234 C; 240 G; 229 T; 0 U; 1 Other;  
 Query Match 99.8%; Score 583; DB 10; Length 933;  
 Best Local Similarity 99.8%; Pred. No. 1.1e-180;  
 Matches 583; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GGGACATTATCCAGAAGATTCTCTAGACGCTATTACAACTCAGCCCTGAAGACCATGA 60  
 DB 350 GGGACATTATCCAGAAGATTCTCTAGACGCTATTACAACTCAGCCCTGAAGACCATGA 409  
 QY 61 TGGGCATAGATGGTCCATCGCTTTTGTCTCAGGGCGGCTTTGTGATGAAACAGACT 120  
 DB 410 TGGGCATAGATGGTCCATCGCTTTTGTCTCAGGGCGGCTTTGTGATGAAACAGACT 469  
 QY 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAACAGAACAA 180  
 DB 470 CNGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAACAGAACAA 529  
 QY 181 CCAGGTTTTTCACTGGCTCTTGAACCTCATGATGATTTCCCATCAGGCGCCATTCAGCA 240  
 DB 530 CCAGGTTTTTCACTGGCTCTTGAACCTCATGATGATTTCCCATCAGGCGCCATTCAGCA 589  
 QY 241 AGTGGTTTGTGATGAAATCTGAATATCCGTGGGACAGGTACCCACCATTCGTCTCCGGCA 300



Db 590 AGTGTTTGTCTGTAATCTGAATATCCGTGGGACAGGTACCCACCATCTGTCTCCGGCA 649  
 Qy 301 CCGCTACGTGTTTCTGCGACGTCGGGAGTCAGGTGTACAATGTCTCCAGAGCGTCC 360  
 Db 650 CCGCTACGTGTTTCTGCGAGCTGGGAGTCAGGTGTACAATGTCTCCAGAGCGTCC 709  
 Qy 361 CATACATTTAACTGAAGACGTGTTTGTGGGCTCTGCTCGAAAGGCTGAACATCAGAT 420  
 Db 710 CATACATTTAACTGAAGACGTGTTTGTGGGCTCTGCTCGAAAGGCTGAACATCAGAT 769  
 Qy 421 TGGAGAGCTCCACCTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTCTCCGATGCC 480  
 Db 770 TGGAGAGCTCCACCTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTCTCCGATGCC 829  
 Qy 481 TCTTCAGGAGGATCGTGGCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGCC 540  
 Db 830 TCTTCAGGAGGATCGTGGCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGCC 889  
 Qy 541 AGGCTCTAGAGAAATCCCGGGGGGAAGATTTGTCGCCCTGTCTGA 584  
 Db 890 AGGCTCTAGAGAAATCCCGGGGGGAAGATTTGTCGCCCTGTCTGA 933

## RESULT 9

AA35710  
 ID AAX35710 standard; cDNA; 1266 BP.

AC AAX35710;

DT 09-JUL-1999 (first entry)

XX cDNA encoding a protein identified by the signal sequence trap method.

XX Signal sequence trap method; SST method; immunisation; inhibition;  
 KW infection; allergy; cancer; regulation; tissue formation; tissue repair;  
 KW activin activity; inhibin activity; chemokine activity;  
 KW cytokine activity; blood coagulation regulation; agonist;  
 KW metabolic disorder; hormonal disorder; immune disorder;  
 KW severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer; wound;  
 KW ss.

XX Homo sapiens.

XX WO9918126-A1.

XX 15-APR-1999.

XX 06-OCT-1998; 98WO-JP004514.

XX 07-OCT-1997; 97JP-00274674.

XX (ONOY ) ONO PHARM CO LTD.

XX Fukushima D, Shibayama S, Tada H;

XX WPI; 1999-277254/23.

XX P-PSDB; AAY02366.

XX Polypeptides identified by the signal sequence trap method from a human  
 PT cDNA library.

XX Claim 4; Page 108-109; 281pp; Japanese.

XX AAX35694-X35747 represent cDNA sequences that encode novel polypeptides  
 CC (AAY02358-84) which are identified from a human placental cDNA library by  
 CC the signal sequence trap (SST) method. The polypeptides have a broad  
 CC range of physiological activity, including immunisation against and  
 CC inhibition of infectious, allergies and cancer; regulation of tissue  
 CC formation and repair; activin/inhibin activity; chemokine/cytokine  
 CC activity; blood coagulation regulation; and receptor/ligand agonist or  
 CC antagonist activity. The polypeptides can be used for prevention and  
 CC treatment of disorders including infections by bacteria, yeasts and  
 CC viruses (including HIV) and protozoa; metabolic and hormonal disorders;

CC immune disorders (including severe combined immunodeficiency (SCID) and  
 CC AIDS; thrombosis; cancer; and traumatic or surgical wounds  
 XX  
 SQ Sequence 1266 BP; 408 A; 270 C; 242 G; 346 T; 0 U; 0 Other;

Query Match 19.6%; Score 114.4; DB 2; Length 1266;

Best Local Similarity 51.8%; Pred. No. 2e-26;

Matches 287; Conservative 0; Mismatches 261; Indels 6; Gaps 1;

Qy 3 GACATATCCAGAAGGATTTCTAGACGTCTATTACAATCTGACCTGAAGACCATGATG 62

Db 640 GATATAATTCACAGGAATACTTAGATAGCTACTATAATTGGACCATTAACACTAATG 699

Qy 63 GGCATAGAAATGGGTCCATCGCTTTTGTCTCAGCGCGGTTTGTGATGAAAACAGACTCA 122

Db 700 GGCATGAATCGGTTGCAACATACCTGTCACATATTTCCATATGTTATGAAAACAGACT 759

Qy 123 GACATGTTCAATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAAACAGAACACC 182

Db 760 GACATGTTTGTCAACACACTGAATATTTAATCAATTAAGTTACTGAAGCCAGATCTGCCCTCC 819

Qy 183 AG-----GTUTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCGCCATTTC 236

Db 820 AGACATAACTATTTTCACTGGTTTACCTAATGCGAGGATATGCACCCCATCGAAAACAAGAT 879

Qy 237 AGCAAGTGGTTTGTTCAGTAAATCTGAATATATCCGTGGGACAGGTACCCACCATTTCTGCTCC 296

Db 880 AGCAAGTGGTACATGCCACAGACCTCTACCCAAGTGAGCGTTATCTCTTCTTCTTCT 939

Qy 297 GGCACCGGCTACGTGTTTCTTGGGACGCTGGGAGTCAGGTGTACAATGTCTCCAAGAGC 356

Db 940 GGAACCTGGTTATGTTTCTTGGAGATCTGCGAGAAAAGATTTTAAAGTTTCTTTAGGT 999

Qy 357 GTCCCATACATTAACATGGAAGAGCGTGTTCGTGGGGCTCTGCTCGAAAGGCTGAACATC 416

Db 1000 ATCCGCGTTTGCATCTGGAAGATGTATGTAGGAGTCTGCTTCCCAAGTTGAGATT 1059

Qy 417 AGATTGGAGGAGCTCCACTCCACGCGGACCTTTTTCAGGGGGGCTTACGCTTCTCCGTA 476

Db 1060 GATCCTGTACCCCTCCCAATGAGTTTGTGTTCAATCACTGGCGAGTCTCTTATTTCGAGC 1119

Qy 477 TGCCTCTTCAGGAGGATCGTGGCTGCGCATTCATCAAGCCTCGGACTCTCTTGGACTAC 536

Db 1120 TGTAAATACAGCCACCTAAATTACCTCTCATCAGTTCCAGCCTAGTGAACATGATAAATAC 1179

Qy 537 TGGCAGGCTCTAGA 550

Db 1180 TGGAAACCAATTACA 1193

## RESULT 10

ABZ56911

ID ABZ56911 standard; DNA; 1268 BP.

XX AC ABZ56911;

XX DT 04-APR-2003 (first entry)

XX DE B3GALT nucleic acid sequence # SEQ ID 3.

XX B3GALT; beta-1,3 galactosyltransferase; p53; cancer; breast; colon;  
 KW kidney; lung; ovary; gene; ds.

XX OS Homo sapiens.

XX PN WO200299044-A2.

XX PD 12-DEC-2002.

XX PF 02-JUN-2002; 2002WO-US017356.

XX PR 05-JUN-2001; 2001US-0296076P.

XX PR 10-OCT-2001; 2001US-0328605P.

PR 15-FEB-2002; 2002US-0357253P.  
 XX (EXEL-) EXELIXIS INC.  
 XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
 XX WPI; 2003-156849/15.  
 XX Identifying p53 pathway modulating agents with B3GALT genes, useful for  
 XX the diagnosis and treatment of disorders associated with defects in the  
 XX p53 pathway, such as cancer of the breast, colon, kidneys, lung and  
 XX ovary.  
 XX Disclosure; Page 42-43; 82pp; English.  
 XX The invention relates to identifying a candidate p53 pathway modulating  
 XX agent in humans that is referred to in the specification as BGALT (beta-  
 XX 1,3 galactosyltransferase). Also disclosed is a method for diagnosing a  
 XX disease in a patient, by contacting a sample with a probe for B3GALT  
 XX expression, and comparing the results with a control, and determining  
 XX whether the results indicate a likelihood of disease. Methods and  
 XX compositions of the invention are useful for the diagnosis and treatment  
 XX of disorders associated with defects in the p53 pathway, such as cancer  
 XX of the breast, colon, kidneys, lung and ovary. The current sequence  
 XX represents a B3GALT nucleic acid sequence referred to in the disclosure  
 XX of the invention  
 XX SQ Sequence 1268 BP; 409 A; 270 C; 242 G; 347 T; 0 U; 0 Other;  
 Query Match 19.6%; Score 114.4; DB 10; Length 1268;  
 Best Local Similarity 51.8%; Pred. No. 2e-26;  
 Matches 287; Conservative 0; Mismatches 261; Indels 6; Gaps 1;  
 QY 3 GACATTATCCAGAGGATTTCCTAGACGTCTATTACATCTGACCTGAAGACCATGATG 62  
 DB 640 GATATAATTCAACAGGAATCTTGTCTCTAGTCTGATCTATATTTGACCATTTAAACACTAATG 699  
 QY 63 GGCATAGATGGTCCATCGCTTTTGTCTCTAGCGGGTGTGTGATGAAACAGACTCA 122  
 DB 700 GGCATGAACTGGTGGTGAACACTGCTCCATATTCATATGTTATGAAACACTGACAGT 759  
 QY 123 GACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAACAGAACACACC 182  
 DB 760 GACATGTTTGTCAACACTGAATATTTAATCAATGTTTACTGAAGCCAGATCTGCCTCCC 819  
 QY 183 AG-----GTTTTCACGTGGCTTCTTGAACACTCAATGATTTCCATCAGCAGCCATTC 236  
 DB 820 AGACATRAACTATTTCACTGGTTTACCTAATGCGAGGATATGCACCCCAATCGAAACAAAGAT 879  
 QY 237 AGCAAGTGGTGTGTCAGTAAATCTGAATATCCGTGGGACAGTACCCACCATCTGTCTCC 296  
 DB 880 AGCAAGTGGTATGCGACACGACCTCTACCAAGTGAAGGTTATCTCTGTTCTTCTTCT 939  
 QY 237 GGCACCGCTACGTGTTTTCTGGGACGTGGGAGTCAAGTGTCAATGTCTTCCAAAGAGC 356  
 DB 940 GGAACGTGTTATGTTTTCTGGAGATCTGCGAGAAAGATTTTTTAAAGTTCTTTAGGT 999  
 QY 357 GTCCATACATTAATCTGGAAGAGTGTGTTGGGCTCTGCTCGAAGGCTGAACATC 416  
 DB 1000 ATCCGCGTGTGCACTTGAAGAGTATATGTAGGGATCTGCTTTCGCAAGTGTGAGAT 1059  
 QY 417 AGATTGGAGGAGCTCCACTCCAGCGGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTA 476  
 DB 1060 GATCTGTACCCCTCCCAATGAGTTTGTGTTCAATCACTGCGAGTCTTATTTGAGC 1119  
 QY 477 TGCCTCTTTGAGGAGGATCGTGCCCTGCCACTTTCATCAAGCCTCGGACTCTCTTGGACTAC 536  
 DB 1120 TGTAAATACAGCCACTAATTAACCTCTCATCAGTTCCAGCCTAGTGAACATGATAAATAC 1179  
 QY 537 TGGCAGGCTCTAGA 550  
 DB 1180 TGAACCATTTTACA 1193

RESULT 11  
 AAD48744  
 ID AAD48744 standard; DNA; 1269 BP.  
 XX AAD48744;  
 AC  
 XX 07-MAR-2003 (first entry)  
 DT Murine beta3GalT2 gene.  
 DE Murine beta3GalT2 gene.  
 XX Murine; transgenic; transgenic animal; muscle coordination disorder;  
 KW UDP-galactose; beta-N-acetylglucosamine beta 1,3-galactosyltransferase;  
 KW beta3GalT2; schizophrenia; depression; atherosclerosis; neuroleptic;  
 KW coronary artery disease; inotropic; gene; ds.  
 XX Mus sp.  
 OS  
 XX Key Location/Qualifiers  
 FH CDS 1..1269  
 FT /product= "Murine beta3GalT2 protein"  
 FT /tag= a  
 FT misc\_feature 292..1119  
 FT /note= "Sequence flanking Neo insert in targeting  
 FT construct"  
 FT /tag= b  
 FT misc\_feature 1120..1237  
 FT /tag= c  
 FT /note= "Sequence deleted in targeting construct"  
 FT 1238..1269  
 FT /tag= d  
 FT /note= "Sequence flanking Neo insert in targeting  
 FT construct"  
 FT  
 XX WO200279424-A2.  
 FN 10-OCT-2002.  
 XX 29-MAR-2002; 2002WO-US009716.  
 XX 29-MAR-2001; 2001US-0280362P.  
 PR 02-OCT-2001; 2001US-0326700P.  
 XX (DELT-) DELTAGEN INC.  
 XX Leviten MW;  
 XX WPI; 2003-067442/06.  
 DR P-PSDB; AAE31623.  
 XX New transgenic mouse comprising a disruption in a beta3GalT2 gene, useful  
 XX for preparing a pharmaceutical composition for treating schizophrenia,  
 XX depression, muscle coordination disorder, atherosclerosis or coronary  
 XX artery disease.  
 XX Example 1; Fig 1; 81pp; English.  
 XX The present invention relates to transgenic animals, compositions and  
 XX methods relating to the characterisation of gene function. The invention  
 XX also relates to transgenic mice comprising a disruption in UDP-galactose:  
 XX beta-N-acetylglucosamine beta 1,3-galactosyltransferase (beta3GalT2)  
 XX gene. The transgenic mice of the invention are useful for preparing  
 XX pharmaceutical compositions for treating schizophrenia, depression, muscle  
 XX coordination disorder, atherosclerosis or coronary artery disease. The  
 XX present sequence is murine beta3GalT2 gene  
 XX Sequence 1269 BP; 400 A; 284 C; 254 G; 331 T; 0 U; 0 Other;  
 Query Match 19.6%; Score 114.4; DB 8; Length 1269;  
 Best Local Similarity 51.8%; Pred. No. 2e-26;  
 Matches 287; Conservative 0; Mismatches 261; Indels 6; Gaps 1;  
 QY 3 GACATTATCCAGAGGATTTCTAGACGTCTATTACATCTGACCTGAAGACCATGATG 62

Db 640 GATATAATTCAGCAGGAATATTTAGATACATATATATCTGACCAATTTAAACACACTAATG 699  
 Qy 63 GGCATAGAAATGGTCCATCGCTTTTGTCTCAGGGCGGTTTGTGATGAAACAGACTCA 122  
 Db 700 GGTATGAATGGTGGCAACACTGTCTCCACATCTCCCTATGTATGAAACGGACAGT 759  
 Qy 123 GACATGTTTCATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAAGAAAACAGAACACC 182  
 Db 760 GACATGTTTGTCAACACAGAAATCTTAATACACAAAGTTACTAAAGCCAGACTGCTCCT 819  
 Qy 183 AG-----GTTTTTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGCAGCCATTTC 236  
 Db 820 AGACATAAATACTATTTTACTGGCTATCTAATGAGAGGATATGACCCGAAACAGAAACAGAC 879  
 Qy 237 AGCAAGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTACCCACCACTTCTGCTCC 296  
 Db 880 AGTAAGTGGTACATGCCACAGACCTTTACCCAGTGGAGCGCTACCTGTCTTCTGCTCA 939  
 Qy 297 GGCACCGCTACGTGTTTCTGGGACGCTGGGAGTCAAGTGTACAATGTCTTCCAAGAGC 356  
 Db 940 GGAACGTGTTATGTTTCTGGGATCTGGCAGAGAAGATATTTAAGGTTTCTTTAGGT 999  
 Qy 357 GTCCCATACATTAACCTGGAAGAGTGTGTTGGGCTCTGCTCGAAAGGCTGAACATC 416  
 Db 1000 ATCCGTCGTTTGCACCTTGGAAAGATGATATGTAGGGATCTGTCTTGGCCAAAGTTGAGAGTT 1059  
 Qy 417 AGATTGGAGGAGCTCCACTCCCGAGCGACCTTTTTCAGGGGCTTACGGTCTTCCCGTA 476  
 Db 1060 GATCCTGTGCCCCCTCCCAATGAGTTCGTTTCAATCACTGGCGAGTTCTTATTTCAAGC 1119  
 Qy 477 TGCCTCTTTCAGGAGGATCGTGGCTTCCCACTTTCATCAAGCCTCGGACTCTCTTGGACTAC 536  
 Db 1120 TGTAAATACAGCCACCTAATTTACCTCTCATCAGTTCCAGCCTAGTGAAGTATGATAAATAC 1179  
 Qy 537 TGGCAGGCTCTAGA 550  
 Db 1180 TGGAAACCATTTACA 1193

RESULT 12

ABZ56913  
 ID ABZ56913 standard; DNA; 1269 BP.

XX AC ABZ56913;

DT 04-APR-2003 (first entry)

XX B3GALT nucleic acid sequence # SEQ ID 5.

XX B3GALT; beta-1,3 galactosyltransferase; p53; cancer; breast; colon;

XX kidney; lung; ovary; gene; ds.

XX Homo sapiens.

XX WO200299044-A2.

PD 12-DEC-2002.

XX 02-JUN-2002; 2002WO-US017356.

XX 05-JUN-2001; 2001US-0296076P.

PR 10-OCT-2001; 2001US-0328605P.

PR 15-FEB-2002; 2002US-0357253P.

XX (EXEL-) EXELIXIS INC.

PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;

XX WPI; 2003-156849/15.

XX Identifying p53 pathway modulating agents with B3GALT genes, useful for  
 PT the diagnosis and treatment of disorders associated with defects in the

PT p53 pathway, such as cancer of the breast, colon, kidneys, lung and  
 PT ovary.

XX Example 5; Page 45; 82pp; English.

CC The invention relates to identifying a candidate p53 pathway modulating  
 CC agent in humans that is referred to in the specification as B3GALT (beta-  
 CC 1,3 galactosyltransferase). Also disclosed is a method for diagnosing a  
 CC disease in a patient, by contacting a sample with a probe for B3GALT  
 CC expression, and comparing the results with a control, and determining  
 CC whether the results indicate a likelihood of disease. Methods and  
 CC compositions of the invention are useful for the diagnosis and treatment  
 CC of disorders associated with defects in the p53 pathway, such as cancer  
 CC of the breast, colon, kidneys, lung and ovary. The current sequence  
 CC represents a B3GALT nucleic acid sequence referred to in an example from  
 CC the invention

XX SQ Sequence 1269 BP; 409 A; 269 C; 243 G; 348 T; 0 U; 0 Other;

Query Match 19.6%; Score 114.4; DB 10; Length 1269;

Best Local Similarity 51.8%; Pred. No. 2e-26;

Matches 287; Conservative 0; Mismatches 261; Indels 6; Gaps 1;

Qy 3 GACATTAATCCAGAGGATTTCTAGACGTCTATTACAACTGACCTCGAAGCAATGATG 62

Db 640 GATATAATTCACAGGAATACTTAGATACGTACTATATTTGACCAATTTAAACACACTAATG 699

Qy 63 GGCATAGAAATGGTCCATCGCTTTTGTCTCAGGGCGGTTTGTGATGAAACAGACTCA 122

Db 700 GGCATGAACCTGGGTGCAACATCTGTCCACATATTTCCATATGTTATGAAACACTGACAGT 759

Qy 123 GACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAAGAAAACAGAAACACC 182

Db 760 GACATGTTTGTCAACACTGATATTTAATCAATGTTTACTGAGCCAGATCTGCTCCTCC 819

Qy 183 AG-----GTTTTTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGCGCAATTC 236

Db 820 AGACATAAATACTATTTCACTGGTTTACCTAATGCGAGGATATGCACCCCAATCGAAACAAAGAT 879

Qy 237 AGCAAGTGGTTTGTGCTAGTAAATCTGAAATATCCGTGGGACAGTACCCACCATTTCTGCTCC 296

Db 880 AGCAAGTGGTATACATGCCACAGACTCTTACCAAGTGAAGCGTTATCTGTCTTCTGTTCT 939

Qy 297 GGCACCGGCTACGTGTTTCTTGGGACGCTGCGAGTCAAGTGTACAATGTCTTCCAAGAGC 356

Db 940 GGAACCTGTTATGTTTCTTGGAGATCTGCGAGAAAGATTTTTTAAAGTTTCTTTAGGT 999

Qy 357 GTCCCATACATTAACCTGGAAGAGCTGTTTGTGGGCTCTGCTCGAAAGGCTGAACATC 416

Db 1000 ATCCGCGGTTTGCACCTTGGAGATGATATGTAGGGATCTGTCTTGGCCAAAGTTGAGAAAT 1059

Qy 417 AGATTGGAGGAGCTCCACTCCCGACCGACCTTTTTCAGGGGCTTACGCTTCTCCGTA 476

Db 1060 GATCCTGTACCCCTCCCAATGAGTTTGTGTTTCAATCACTGCGGAGTCTCTTATTCGAGC 1119

Qy 477 TGCCTCTTCAGGAGGATCGTGGCTTCCACTTTCATCAAGCCTCGGACTCTCTTGGACTAC 536

Db 1120 TGTAAATACAGCCACCTAATTTACCTCTCATCAGTTCCAGCCTAGTGAAGTATGATAAATAC 1179

Qy 537 TGGCAGGCTCTAGA 550

Db 1180 TGGAAACCATTTACA 1193

RESULT 13

ADM66890

ID ADM66890 standard; DNA; 1473 BP.

XX AC ADM66890;

XX 03-JUN-2004 (first entry)

DT Murine adipocyte specific beta 1,3 galactosyltransferase DNA SeqID 23.

XX murine; mouse; adipocyte specific; gene; ds; adipose tissue;  
KW anti-obesity; high mobility group I-C protein; HMGI-C; obesity; leptin;  
KW ob; diabetes; adipogenesis; hypertension; cardiovascular disease;  
KW anorectic; antidiabetic; hypotensive; beta 1,3 galactosyltransferase.  
XX  
OS Mus musculus.  
XX  
XX WO2004011618-A2.  
XX  
XX 05-FEB-2004.  
XX  
XX 29-JUL-2003; 2003WO-US023684.  
XX  
XX 29-JUL-2002; 2002US-0398785P.  
PR 12-JUN-2003; 2003US-0478206P.  
XX  
XX (HMGR-) HMGNE INC.  
XX  
XX Chada K, Chouinard R, Ashar H, Sayed AMD;  
XX  
XX WPI; 2004-143846/14.  
DR P-PSDB; ADM67170.  
XX  
XX Identifying adipocyte specific genes, useful for treating obesity or  
PT diabetes, and for identifying drug targets, by differential gene  
PT expression analysis between adipose tissue or stromal vascular tissue of  
PT mice of different genotypes.  
XX  
XX Claim 11; SEQ ID NO 23; 91pp; English.  
XX  
XX This invention relates to a novel method for identifying genes that are  
CC over-expressed in adipose tissue and as such it provides targets for anti-  
CC obesity pharmaceutical compositions. Specifically, it refers to a high  
CC mobility group I-C protein (HMGI-C) that is associated with obesity and  
CC is epistatic to leptin, furthermore, it refers to the ob gene where an  
CC autosomal recessive trait is linked to obesity and diabetes. The present  
CC invention describes performing differential gene expression analysis  
CC between the white adipose tissue (WAT) or stromal vascular tissue (SVT)  
CC of any two different mice selected from a group consisting of wild-type,  
CC HMGI-C<sup>-/-</sup>, ob/ob, or HMGI-C<sup>-/-</sup> ob/ob genotype mice. Accordingly, using  
CC this method novel nucleotides and the encoded proteins thereof were  
CC identified that are adipocyte specific, and as such can be used for  
CC preventing adipogenesis, diagnosing and treating diabetes, obesity,  
CC hypertension and cardiovascular disease, as well as screening for  
CC compounds that can modulate or prevent adipogenesis and treat diabetes or  
CC obesity. These compositions exhibit anorectic, antidiabetic and  
CC hypotensive activities. This polynucleotide sequence is a murine  
CC adipocyte specific DNA sequence of the invention.  
XX  
XX Sequence 1473 BP; 470 A; 315 C; 288 G; 400 T; 0 U; 0 Other;  
SQ  
Query Match  
Best Local Similarity 19.6%; Score 114.4; DB 12; Length 1473;  
Matches 287; Conservative 0; Mismatches 261; Indels 6; Gaps 1;  
  
QY 3 GACATATCCAGAAGATTCTAGACGTCTATTACAACTGACCTGAAGACCATGATG 62  
Db 702 GATATAATTCAGCAGGATATTAGATACATACATCTATCTGACCATTAACACTAATG 761  
QY 53 GGCATAGATGGTCCATCGCTTTTGTCTCAGGGGGGTTTGTGATGAAACAGACTCA 122  
Db 712 GGTATGAATCGGTTTGAACACTGTCTCCATACATCTCCCTATGTTATGAAACCGACAGT 821  
QY 183 GACATGTTTCATCAATGTTGACTATCTGACTGAACCTCTTGAAGAAAAACAGAACACC 182  
Db 812 GACATGTTTGTCAACACAGATACATCTATACACAGATTACTAAAGCCAGACTGCCTCT 881  
QY 183 AG-----GTTTTTCAGTGGCTTCTTGAACCTCAATGATGTTCCCATCAGGAGCCATTC 236  
Db 812 AGACATAACTATTTTACTGGCTTCTAATGATGAGGATATGCACCGAACAGAAACAAAGAC 941  
QY 217 AGCAAGTGGTTGTTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATCTGCTCC 296

Db 942 AGTAAGTGGTACATGCCACGACCTTTTACCAAGTGAGCGCTACCCCTGCTCTGCTCA 1001  
QY 297 GGCACCGGCTACGTGTTTCTGGCGAGTGGCGAGTCAGGTGTACAAATGTCTCCAGAGC 356  
Db 1002 GGAACCTGTTATGTTTCTGGGATCTGGCAGAGAGATATTTAAGGTTTCTTTAGGT 1061  
QY 357 GTCCCATACATTAAACTGGAAGACGTTTGTGGGCTCTGCCCTCGAAAGGCTGAACATC 416  
Db 1062 ATCCGTCGTTTGCACCTTGGGAAGATGTATATGTAGGGAATCTGTTGCCAAGTTGAGAGTT 1121  
QY 417 AGATTGGAGGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTA 476  
Db 1122 GATCTCTGCCCCCTCCCAATGAGTTGCTGTTCAATCACTGGCGAGTTCTTATTCAAGC 1181  
QY 477 TGCCTCTTTCAGGAGGATCGTGGGCTGCGCACTTCAATCAAGCCTCGGACTCTCTTGGACTAC 536  
Db 1182 TGTAAATACAGCCACCTAAATTAACCTCTCATCAGTTCAGGCTAGTGAACCTGATAAATAC 1241  
QY 537 TGGCAGGCTCTAGA 550  
Db 1242 TGGAAACCAATTTACA 1255  
  
RESULT 14  
AAX35711  
ID AAX35711 standard; cDNA; 2420 BP.  
XX  
XX AAX35711;  
XX  
XX 09-JUL-1999 (first entry)  
XX  
XX cDNA encoding a protein identified by the signal sequence trap method.  
DE  
XX  
XX Signal sequence trap method; SST method; immunisation; inhibition;  
KW infection; allergy; cancer; regulation; tissue formation; tissue repair;  
KW activin activity; inhibin activity; chemokine activity;  
KW cytokine activity; blood coagulation regulation; agonist; antagonist;  
KW metabolic disorder; hormonal disorder; immune disorder;  
KW severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer; wound;  
XX ss.  
XX  
XX Homo sapiens.  
XX  
XX WO9918126-A1.  
XX  
XX 15-APR-1999.  
XX  
XX 06-OCT-1998; 98WO-JP004514.  
XX  
XX 07-OCT-1997; 97JP-00274674.  
XX  
XX (ONOY ) ONO PHARM CO LTD.  
XX  
XX Fukushima D, Shibayama S, Tada H;  
XX  
XX WPI; 1999-277254/23.  
XX  
XX P-PSDB; AAY02366.  
XX  
XX Polypeptides identified by the signal sequence trap method from a human  
PT cDNA library.  
XX  
XX Claim 5; Page 110-113; 281pp; Japanese.  
XX  
XX AAX35694-X35747 represent cDNA sequences that encode novel polypeptides  
CC (AAY02358-84) which are identified from a human placental cDNA library by  
CC the signal sequence trap (SST) method. The polypeptides have a broad  
CC range of physiological activity, including immunisation against and  
CC inhibition of infections, allergies and cancer; regulation of tissue  
CC formation and repair; activin/inhibin activity; chemokine/cytokine  
CC activity; blood coagulation regulation; and receptor/ligand agonist or  
CC antagonist activity. The polypeptides can be used for prevention and  
CC treatment of disorders including infections by bacteria, yeasts and

CC	viruses (including HIV) and protozoa; metabolic and hormonal disorders;
CC	immune disorders (including severe combined immunodeficiency (SCID) and
CC	AIDS; thrombosis; cancer; and traumatic or surgical wounds
XX	
SQ	Sequence 2420 BP; 853 A; 458 C; 451 G; 658 T; 0 U; 0 Other;
Query Match	19.6%; Score 114.4; DB 2; Length 2420;
Best Local Similarity	51.8%; Pred. No. 2.9e-26;
Matches 287; Conservative	0; Mismatches 261; Indels 6; Gaps 1;
QY	3 GACATTATCCAGAGGATTTCTTAGAGCTCTATTACAACTCTGACCCCTGGAAGACCATCATG 62
DB	1337 GATATAATTCAACAGGAATACCTTAGATACGTACTATAATTGACCATTAATAAACTAATG 1396
QY	63 GGCATGAATGGGTCCATCGCTTTGTCTCCAGCGGGCTTTGTGATGAAAAACAGACTCA 122
DB	1397 GGCATGAATGGGTGCAACATACGTGCCATATTCATATGTTATGAAAACTGACAGT 1456
QY	123 GACATGTTCAATATGTTGACTATCTGACTGAACTGCTTCTGGAAGAAAAACAGAAACAC 189
DB	1457 GACATGTTTGTCAACACTGAAATATTTAATCAATAAGTTACTTGAAGCCAGATCTGCCTCCC 1516
QY	183 AG-----GTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGCGAGCCATT 236
DB	1517 AGACATAACTATTTTCACTGGTTACTTAATGCGAGGATATGCACCCATCGAAACAAAGAT 1576
QY	237 AGCAAGTGGTTTTGTTCAGTAAATCTGAAATATCCGTGGGACAGGTACCCACCATTCGTCTCC 296
DB	1577 AGCAAGTGGTACATGCCACAGACCTCTACCCAAAGTCAGCGTTATCTGTCTTCTGTTCT 1636
QY	297 GGCACCGGCTACGTTTCTTGGCGAGCGTGGCGAGTCAGGTGTACAATGTCTCCAGAGC 356
DB	1637 GGAACCTGTTATGTTTTTCTCGAGATCTGGCAGAAAAGATTTTAAAGTTTCTTTAGGT 1696
QY	357 GTCCCATACATTAACCTGGAAGACGTTGTTGCGGCTCTGCCTCGAAAGCGCTGAACATC 416
DB	1697 ATCCGCGTTTGCACTTTGGAAGATGTATATGTAGGATCTGTCTTGCCAAAGTTGAGAATT 1756
QY	417 AGATTGAGGAGTCCCACTCCACGCCACCTTTTTTCCAGGGGCTTACGCTTCTCCGTA 476
DB	1757 GATCCTGTACCCCTCCCAATGAGTTGTGTTCAATCACTGGCGAGTCTCTTATTCGAGC 1816
QY	477 TGCTCTTTCAGAGGATCGTGGCGTCCCACTTCAATCAAGCGCTGGACTCTCTTTGGACTAC 536
DB	1817 TGTAAATACAGCCACCTAAATTACCTCTCATCAGTTCAGGCTAGTGAACCTGATAAAATAC 1876
QY	537 TGGCAGGCTCTAGA 550
DB	1877 TGAACCAATTTACA 1890
RESULT 15	
ADM66891	standard; DNA; 3054 BP.
XX	
AC	ADM66891;
XX	
DT	03-JUN-2004 (first entry)
XX	
DE	Human adipocyte specific beta 1.3 galactosyltransferase DNA SeqID 24.
XX	
KW	human; adipocyte specific; gene; ds; adipose tissue; anti-obesity;
KW	high mobility group I-C protein; HMG1-C; obesity; leptin; ob; diabetes;
KW	adipogenesis; hypertension; cardiovascular disease; anorectic;
KW	antidiabetic; hypotensive; beta 1,3 galactosyltransferase.
XX	
OS	Homo sapiens.
XX	
PN	W02004011618-A2.
XX	
PD	05-FEB-2004.
XX	
PF	29-JUL-2003; 2003MO-US023684.

Db	1949	GATCCTGTACCCCTCCCAATGAGTTTGTGTTCAATCACTGGCGAGTCTCTTATTGAGC	2008
Qy	477	TGCCTCTTCAGGAGGATCGTGGCCCTGCCACTTCATCAGCCCTCGGACTCTCTTGGACTAC	536
Db	2009	TGTAATACAGGCCACCTTAATTACCTCTCATCAGTTCCAGCCCTAGTGAAGTATAAAATAC	2068
Qy	537	TGGCAGGCTCTAGA	550
Db	2069	TGGAACCATTTACA	2082

Search completed: September 22, 2005, 12:38:54  
Job time : 351.351 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2005, 12:09:12 ; Search time 2575.06 Seconds  
(without alignments)  
8632.613 Million cell updates/sec

Title: US-10-777-828-8\_COPY\_428\_1011

Perfect score: 584

Sequence: 1 gggacattatccagaaggat.....aagatttcgcctgtctga 584

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

BST:\*

1: gb\_est1:\*

2: gb\_est2:\*

3: gb\_hic:\*

4: gb\_est3:\*

5: gb\_est4:\*

6: gb\_est5:\*

7: gb\_est6:\*

8: gb\_ges1:\*

9: gb\_ges2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	584	100.0	933	9 AY419460	AY419460 Homo sapi
2	419.4	71.8	780	9 AY419461	AY419461 Pan trogl
3	369.6	63.3	434	2 BF748181	BF748181 RC3-BN041
4	368	63.0	5069	3 BC047347	BC047347 Mus muscu
5	363.2	62.2	927	9 AY419462	AY419462 Mus muscu
6	282.4	48.4	766	9 CE132434	CE132434 tigr-ges-
7	184.2	31.5	629	7 CR564113	CR564113 CR564113
8	173	29.6	493	1 AJ507051	AJ507051 AJ507051
9	153.8	26.3	492	5 BP761359	BP761359 BP761359
10	148.2	25.4	626	5 BX302370	BX302370 BX302370
11	131	22.4	901	4 BJ740020	BJ740020 BJ740020
12	122.8	21.0	981	9 AY401366	AY401366 Mus muscu
13	122.8	21.0	2705	3 AK035215	AK035215 Mus muscu
14	121.2	20.8	685	5 BU133295	BU133295 603122104
15	121.2	20.8	842	5 BU350312	BU350312 603527845
16	120	20.5	447	6 CB546569	CB546569 AMGNNUC:N
17	120	20.5	1101	9 CNS05ET1	AL330238 Tetraodon
18	119	20.4	1903	3 AK088407	AK088407 Mus muscu
19	117.8	20.2	879	5 BQ95238	BQ95238 AGENCOURT
20	117.4	20.1	996	9 AY402383	AY402383 Mus muscu
21	117.4	20.1	1514	3 AK003837	AK003837 Mus muscu
22	116	19.9	596	4 BU501651	BU501651 BU501651
23	116	19.9	644	6 CA057816	CA057816 ssalrgb50
24	114.6	19.6	697	6 CB233493	CB233493 AGENCOURT

25	114.4	19.6	582	5 BP374190	BP374190 BP374190
26	114.4	19.6	1238	9 AY420414	AY420414 Homo sapi
27	114	19.5	444	4 BM403751	BM403751 zam6155 Z
c 28	114	19.5	574	2 BE652985	BE652985 UI-M-AMI-
29	113.4	19.4	810	7 CF749272	CF749272 UI-M-HJO-
30	112.8	19.3	682	6 BY704097	BY704097 BY704097
31	112.8	19.3	787	7 CK475539	CK475539 AGENCOURT
32	112.8	19.3	994	7 CF585163	CF585163 AGENCOURT
33	112.8	19.3	1238	9 AY420416	AY420416 Mus muscu
34	112.8	19.3	2681	3 AK036141	AK036141 Mus muscu
35	112.8	19.3	2690	3 AK083168	AK083168 Mus muscu
36	112.8	19.3	4427	3 AK034371	AK034371 Mus muscu
37	111.6	19.1	981	9 AY401365	AY401365 Pan trogl
c 38	111.2	19.0	632	4 BJ632719	BJ632719 BJ632719
39	110.6	18.9	981	9 AY401364	AY401364 Homo sapi
40	109.6	18.8	729	7 CF113712	CF113712 Shultzomi
c 41	108.6	18.6	632	9 CC554157	CC554157 CH240 460
42	108.4	18.6	965	9 CNS05IB2	AL338663 Tetraodon
43	108.2	18.5	831	7 CO398909	CO398909 AGENCOURT
44	107.8	18.5	1020	7 CN646142	CN646142 ILLUMIGEN
45	106.2	18.2	712	1 AL042887	AL042887 DKF2p434I

ALIGNMENTS

RESULT 1  
LOCUS AY419460 933 bp DNA linear GSS 17-DEC-2003  
DEFINITION Homo sapiens B3GALT5 gene, VIRTUAL TRANSCRIPT, partial sequence,  
genomic survey sequence.  
ACCESSION AY419460  
VERSION AY419460.1 GI:39775417  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 933)  
AUTHORS Clark,A.G., Nielson,R., Thomas,P., Kejarawal,A.,  
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 933)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,  
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT These sequences were made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
source  
Location/Qualifiers  
1..933  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
<1..>933  
/gene="B3GALT5"  
/locus\_tag="HCN6898"

Query Match 100.0%; Score 584; DB 9; Length 933;  
Best Local Similarity 100.0%; Pred. No. 6.1e-165;  
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGGACATTATCCAGAAGATTTCCTAGACGTCGTATTACATCTGACCTCGAGACCATGA 60  
|||||

```
Db 350 GGGACATTATCCAGAAGGATTTCCTAGAGTCTATTACAACTGACCTCGAAGACCATGA 409
Qy 41 TGGGCATAGATGGGTCCATCGCTTTTGTCTCAGCGCGGTTTGTGATGAACACAGACT 120
Db 410 TGGGCATAGATGGGTCCATCGCTTTTGTCTCAGCGCGGTTTGTGATGAACACAGACT 469
Qy 181 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTCTTCTGAAGAAAAACAGAACAA 180
Db 470 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTCTTCTGAAGAAAAACAGAACAA 529
Qy 181 CCAGGTTTTTTCATCGGCTTCTTGAAGCTCAATGAGTTTCCCATCAGGCAGCCATTCAGCA 240
Db 530 CCAGGTTTTTTCATCGGCTTCTTGAAGCTCAATGAGTTTCCCATCAGGCAGCCATTCAGCA 589
Qy 241 AGTGGTTTGTGATGAATCTGAATATCCGTGGGACAGGTACCCACCATTCCTGCCGCA 300
Db 590 AGTGGTTTGTGATGAATCTGAATATCCGTGGGACAGGTACCCACCATTCCTGCCGCA 649
Qy 301 CCGGCTACGTGTTTCTGGCGAGTGGCGAGTCAAGTGTGACAACTCTCCAAAGAGCGTCC 360
Db 650 CCGGCTACGTGTTTCTGGCGAGTGGCGAGTCAAGTGTGACAACTCTCCAAAGAGCGTCC 709
Qy 361 CATACATTAAACTGGAAGACGTTTGTGGGCTCTGCCTCGAAAGGCTGAACATCAGAT 420
Db 710 CATACATTAAACTGGAAGACGTTTGTGGGCTCTGCCTCGAAAGGCTGAACATCAGAT 769
Qy 421 TGGAGAGCTCCACTCCAGCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC 480
Db 770 TGGAGAGCTCCACTCCAGCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC 829
Qy 481 TCTTCAGAGGAGTCTGGCTGCCACTTTCATCAAGCCTCGGACTCTCTTGGACTACTGGC 540
Db 830 TCTTCAGAGGAGTCTGGCTGCCACTTTCATCAAGCCTCGGACTCTCTTGGACTACTGGC 889
Qy 541 AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCCGCTGTCTGA 584
Db 890 AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCCGCTGTCTGA 933

RESULT 2
AY419461 780 bp DNA linear GSS 17-DEC-2003
LOCUS Pan troglodytes B3GALT5 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY419461
VERSION AY419461.1 GI:39775418
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
1 (bases 1 to 780)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 1 (bases 1 to 780)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
AUTHORS Direct Submission
TITLE Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
JOURNAL Rockville, MD 20850, USA
PUBMED
COMMENT These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES Location/Qualifiers
1..780
source /organism="Pan troglodytes"
```

```
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>780
/gene="B3GALT5"
/locus_tag="HCM6898"

ORIGIN
Query Match 71.8%; Score 419.4; DB 9; Length 780;
Best Local Similarity 98.1%; Pred. No. 3.3e-115; Indels 0; Gaps 0;
Matches 423; Conservative 0; Mismatches 8;

Qy 1 GGGACATTATCCAGAAGGATTTCCTAGACGCTCTATTACAATCTGACCTCGAAGACCATGA 60
Db 350 GGGACATTATCCAGAAGGATTTCCTAGACGCTCTATTACAATCTGACCTCGAAGACCATGA 409
Qy 61 TGGGCATAGATGGGTCCATCGCTTTTGTCTCAGCGCGGTTTGTGATGAACACAGACT 120
Db 410 TGGGCATAGATGGGTCCATCGCTTTTGTCTCAGCGCGGTTTGTGATGAACACAGACT 469
Qy 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTCTTCTGAAGAAAAACAGAACAA 180
Db 470 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTCTTCTGAAGAAAAACAGAACAA 529
Qy 181 CCAGGTTTTTTCATCGGCTTCTTGAAGCTCAATGAGTTTCCCATCAGGCAGCCATTCAGCA 240
Db 530 CCAGGTTTTTTCATCGGCTTCTTGAAGCTCAATGAGTTTCCCATCAGGCAGCCATTCAGTA 589
Qy 241 AGTGGTTTGTGATGAATCTGAATATCCGTGGGACAGGTACCCACCATTCCTGCCGCA 300
Db 590 AGTGGTTTGTGATGAATCTGAATATCCGTGGGACAGGTACCCACCATTCCTGCCGCA 649
Qy 301 CCGGCTACGTGTTTCTGGCGAGTGGCGAGTCAAGTGTGACAACTCTCCAAAGAGCGTCC 360
Db 650 CCGGCTACGTGTTTCTGGCGAGTGGCGAGTCAAGTGTGACAACTCTCCAAAGAGCGTCC 709
Qy 361 CATACATTAAACTGGAAGACGTTTGTGGGCTCTGCCTCGAAAGGCTGAACATCAGAT 420
Db 710 CATACATTAAACTGGAAGACGTTTGTGGGCTCTGCCTCGAAAGGCTGAACATCAGAT 769
Qy 421 TGGAGAGGCTC 431
Db 770 TGGAGGAGCTC 780

RESULT 3
BF748181/c
LOCUS BF748181
DEFINITION RC3-BN0411-021000-021-c04 BN0411 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF748181
VERSION BF748181.1 GI:12074857
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 434)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Ngai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsuoka,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
```



Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tbl=RC3&t2=RC3-BN0411-021000-021-c04&t3=2000-10-02&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 24  
 High quality sequence stop: 432.

## FEATURES

Location/Qualifiers  
 1..434  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="BN0411"  
 /note="Organ: breast normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 63.3%; Score 369.6; DB 2; Length 434;  
 Best Local Similarity 98.9%; Pred. No. 3.2e-100;  
 Matches 372; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 209 CAATGAGTTTCCCATCAGCAGCCATTTCAGCAAGTGTTGTGTCAGTAATCTGAATATCC 268  
 DB 434 CAATGAGTTTCCCATCAGCAGCCATTTCAGCAAGTGTTGTGTCAGTAATCTGAATATCC 375  
 QY 269 GTGGACAGGTACCCACCACTTCGTCCGGCCAGCGTACGTTTTCGTGGCAGGTGC 328  
 DB 374 GTGGACAGGTACCCACCACTTCGTCCGGCCAGCGTACGTTTTCGTGGCAGGTGC 315  
 QY 329 GAGTCAGGTGTCAATGTCTCCAAAGACGCTCCCATACATTAATTCGAAGACGTGTTGT 388  
 DB 314 GAGTCAGGTGTCAATGTCTCCAAAGACGCTCCCATACATTAATTCGAAGACGTGTTGT 255  
 QY 389 GGGGCTTCCTCGAAGCTGAACATCAGATGGAGGAGTCCACTCCAGCCGACCTT 448  
 DB 254 GGGGCTTCCTCGAAGCTGAACATCAGATGGAGGAGTCCACTCCAGCCGACCTT 195  
 QY 449 TTTTCCAGGGGCTTACGCTTCCTCGTATGCTCTTCAGGAGATCGTGGCTGCCACTT 508  
 DB 194 TTTTCCAGGGGCTTACGCTTCCTCGTATGCTCTTCAGGAGATCGTGGCTGCCACTT 135  
 QY 509 CATCAAGCTTCGAGCTCTCTTGAGCTACTCGGAGCTCTAGAGAATTCCTCCGGGGGAAGA 568  
 DB 134 CATCAAGCTTCGAGCTCTCTTGAGCTACTCGGAGCTCTAGAGAATTCCTCCGGGGGAAGA 75  
 QY 569 TTGTCGCTGTCTGA 594  
 DB 74 TTGTCGCTGTCTGA 59

## RESULT 4

BC047347  
 LOCUS BC047347 5069 bp mRNA linear HTC 02-JAN-2004  
 DEFINITION Mus musculus UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 5, mRNA (cdna clone IMAGE:5365404).  
 ACCESSION BC047347  
 VERSION BC047347.1 GI:28611171  
 KEYWORDS HTC.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 5069)  
 AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,

Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.P., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.P., Casavant,T.L., Schetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S., Carninci,P., Frange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickinson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932  
 2 (bases 1 to 5069)  
 Direct Submission  
 Strausberg,R.  
 Submitted (28-FEB-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: http://mgc.nci.nih.gov  
 Contact: MGC help desk  
 Email: cgapbs@mail.nih.gov  
 Tissue Procurement: The Cepko Laboratory  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
 Center code: BCM-HGSC  
 Web site: http://www.hgsc.bcm.tmc.edu/cdna/  
 Contact: amg@bcm.tmc.edu  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Series: IRAP Plate: 53 Row: j Column: 9  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 22122432  
 This clone has the following problem: no 5' EST match.

## FEATURES

## source

Location/Qualifiers  
 1..5069  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5365404"  
 /tissue\_type="Eye, retina, mouse strain C57Bl/6"  
 /clone\_lib="NIH MGC\_94"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"

## ORIGIN

Query Match 63.0%; Score 368; DB 3; Length 5069;  
 Best Local Similarity 76.9%; Pred. No. 2.1e-99;  
 Matches 449; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 1 GGGACATTATCAGAGGATTTCCTAGACGCTTATACAACTGACCTCCCTGAAGACCATGA 60  
 DB 935 GCGACATTATCAGAGGACTTCAAGGATGCCTACTTCAACCTGACCTCCCTGAAGACCATGA 994  
 QY 61 TGGGATAGATGGTCCCATCGCTTTCTCTCCCTCAGCGCGCTTTCTGTGATGAAGACACT 120  
 DB 995 TGGGATAGATGGTCTTACCACCTTTTCTCTCTCAGACAGCTTACGTGATGAAGACGACT 1054  
 QY 121 CAGACATGTTCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAAAGACAA 180

Query Match	62.2%;	Score 363.2;	DB 9;	Length 927;
Best Local Similarity	76.4%;	Pred. No. 3.4e-98;		
Matches 446;	Conservative 0;	Mismatches 138;	Indels 0;	Gaps 0;

QY	1	GGGACATTATCCAGAGGAGTTCCTAGACGTCATTACAAATCTGACCCCTGAAGACCATGA	60
DB	344	GGGACATTATCCAGAGGAGTTCAGAGTGCCTACTTCAACCTGACCCCTGAAACCATGA	403
QY	61	TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGCGGGGTTTGTGATGAAACAGACT	120
DB	404	TGGGTATGAAATGGGTCTACCACTTTTGTCTCAGACAGCTTACGTGATGAAACCGACT	463
QY	121	CAGACATGTTTCATCATGTTGACTATCTGACTGAACTGCTTCTGAGAGAAAAACAGACAA	180
DB	464	CTGACATGTTTGTGAATGTGGCTATCTGACGGAACTGCTGCTTAAAGAAAAACAAACGA	523
QY	181	CCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTTCCATCAGCGACGCCATTTCAGCA	240
DB	524	CCAGGTTCTTCCAGGCTACATAAAGCCCCACGACTTTCCCATCCGGCAGAAGTTCACAA	583
QY	241	AGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTATCCACCAATTCGTCTCGGCA	300
DB	584	AGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTATCCACCACTTTTGTCTGCTGA	643
QY	301	CCGGCTTACGTTGTTTCTGCGCAGCTGCGCAGTCAGGTGTACAATGTCTCCAAGAGCGTCC	360
DB	644	CTGGTTATGTCTTTTCCAGCGAGCTGGCGATCCAGTATACAACGTCCTAGAGAGCGTTC	703
QY	361	CATACATTAACTGGGAAGACGTTTGTGGGGCTCTGCGCTCGAAAGGCTGAACATCAGAT	420
DB	704	CGTTTCATCAAGCTGGAGGATGTGTTTGTGGGCTCTGCTTGGCCAAGCTAAAGATCCGGC	763
QY	421	TGGAGAGCTCCACTCCAGCCGACCTTTTTCAGGGGGCTTACGCTTCTCCGTATGCC	480
DB	764	CGGAGAGCTGCAACACCAACAGACCTTCTTCCTCGCGGTTTACGCTTTTCCGTGTGCC	823
QY	481	TCTTTCAGGAGGATCGTGGCTGCGCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGC	540
DB	824	GCCTTCAGAAAATTGTGGCATGCATTTTATGAAGCCCGCAGGACCTGCTCACTTACTGGC	883
QY	541	AGGCTCTAGAGAAATCCCGGGGGGAAGATGTCCGCTGTGTGA	584
DB	884	AAGCACTGGAGAACTCGAAGAAACAGGACTGCGCTGTGTGTGA	927

RESULT 6  
 CEI32434/c  
 LOCUS  
 DEFINITION  
 CEI32434 766 bp DNA linear GSS 25-SEP-2003  
 tigr-gss-dog-17000326239815 Dog Library Canis familiaris genomic,  
 genomic survey sequence.  
 ACCESSION  
 CEI32434  
 VERSION  
 CEI32434.1 GI:35233692  
 KEYWORDS  
 GSS.  
 SOURCE  
 Canis familiaris (dog)  
 ORGANISM  
 Canis familiaris  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 REFERENCE  
 1 (bases 1 to 766)  
 AUTHORS  
 Kirkness,E.F., Batina V., Halpern,A.L., Levy,S., Remington,K.,  
 Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and  
 Venter,J.C.  
 TITLE  
 The dog genome: survey sequencing and comparative analysis  
 JOURNAL  
 Science 301 (5641), 1898-1903 (2003)  
 MEDLINE  
 22875432  
 PubMed  
 14512627  
 COMMENT  
 Contact: Kirkness EF  
 The Institute for Genomic Research  
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
 Rockville, MD 20850, USA  
 Tel: 301-838-0200  
 Fax: 301-838-0208  
 Email: ekirknes@tigr.org  
 Class: shotgun.

```
FEATURES
  source
    Location/Qualifiers
      1..766
        /organism="Canis familiaris"
        /mol_type="genomic DNA"
        /strain="Standard Poodle"
        /db_xref="taxon:9615"
        /clone_lib="Dog Library"
        /notes="Site 1: BstXI; Libraries were prepared from peripheral blood"

ORIGIN
  Query Match      48.4%; Score 282.4; DB 9; Length 766;
  Best Local Similarity 81.2%; Pred. No. 8.7e-74;
  Matches 328; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1 GGGACATATTCAGAGAGGATTTCTAGAGCTCTATTACAACTCTGACCTGAAGACCATGA 60
DB 415 GCGACATATTCAGAGAGGACTTTGTGAGGCTTACTTCAATTTGACCTGAAGACCATGA 356
QY 61 TGGGCGATAGATGGGTCCATCGCTTTTGTCTCAGCGCGGTTTGTGATGAAGAAACAGACT 120
DB 355 TGGGTATAGATGGATCCACCGCTTCTGTCTCAGCGGAGCTTTGTGATGAAGAGCGACT 296
QY 121 CAGACATGTTTCATCATATGTTGATCTGACTGAAGTCTTCTGAAGAAACAGAACAA 180
DB 295 CGGACATGTTTGTCAAGCTCTACTGACCGAGCTGCTTCTCAAGAAACAGAACCA 236
QY 181 CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCCATCAGCGGACCATTCAGCA 240
DB 235 CTCGGTTTTTCAACCGGCTTCTTAAACTGACAGTTTCCCATTAGGAGCAAGGCCAAC 176
QY 241 AGTGGTTTGTGATGAATCTGAATATCCGTGGGACAGGTACCACCATTCCTCGCGCA 300
DB 175 AGTGGTTTGTGATGAATATACGAATACCATGGGATAAGTACCACCGCTTTTGTCTGGGCA 116
QY 301 CCGGCTACGTGTTTCTGGGAGCGTGGGAGTGTGACATGTCTCTCAAGAGCGTCC 360
DB 115 CCGGCTATGTGTTTCCAGTGTATGTTGCAAGTCAGGTGTACACGCTCTGACAGCGTCC 56
QY 361 CATACATTAATCGAAGACGTGTTTGTGGGCTCTGCGCTCGAA 404
DB 55 CGTTCAATTAAGTCGAGACGCTTTGTGGGACTCGCGGCGCA 12

RESULT 7
CR564113 629 bp mRNA linear EST 19-JUL-2004
LOCUS CR564113 XGC-tailbud-head Xenopus tropicalis cDNA clone Thda01ln08
DEFINITION CR564113 5', mRNA sequence.
ACCESSION CR564113
VERSION CR564113.1 GI:50394190
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
REFERENCE 1 (bases 1 to 629)
AUTHORS Croning,M.D.R., Ashurst,J.L., Taylor,R., Garrett,N. and Rogers,J.
TITLE Sanger Xenopus tropicalis EST project 2001 (2004)
JOURNAL Sanger Xenopus tropicalis EST project 2001 (2004)
COMMENT Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE ID: Thda01ln08.plkSP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.
Seq primer: SP6.
Location/Qualifiers
  1..629
  /organism="Xenopus tropicalis"

FEATURES
  source
    Location/Qualifiers
      1..766
        /organism="Canis familiaris"
        /mol_type="genomic DNA"
        /strain="Standard Poodle"
        /db_xref="taxon:9615"
        /clone_lib="Dog Library"
        /notes="Site 1: BstXI; Libraries were prepared from peripheral blood"

ORIGIN
  Query Match      31.5%; Score 184.2; DB 7; Length 629;
  Best Local Similarity 65.1%; Pred. No. 4e-44;
  Matches 287; Conservative 0; Mismatches 153; Indels 1; Gaps 1;

QY 102 TTTGTGATGAAGACGACTCAGACATG-TTCATCATGTTGACTATCTGACTGAACGTCT 160
DB 1 TTTGTGATGAAGACGACTCAGACATGTTTCGTTAACACCTTCTACCTGGTCCAACTGCT 60
QY 161 TCTGAAGAAAAACAGAAACACAGGTTTTTCACTGGCTTCTTGAAACTCAATCAGTTTCC 220
DB 61 GCGAAGAAAAACAGGCTCTTCTAATTTTTTACTGGTTTCTCAAACTGAACGAGTACCC 120
QY 221 CATCAGGACGACCATTCAGCAAGTGTCTCAGTAAATCTGAATATCCGTGGGACAGGTA 280
DB 121 GATAAGGAATATCTTCAGCAAGTGTGTACGCCAGTAAAGGGAATACCCAGGGGCCAAGTA 180
QY 281 CCCACCATTCCTCTCGGCACCGGTACGTGTTTCTGGGACGTGGCGAGTCAGGTGTA 340
DB 181 CCTCCATTTTGTTCGGGAGCTGATACGTCTTTCTGTAGACGTGCGCCAAAAAGATCCA 240
QY 341 CAATGTCTCCAAAGAGCGTCCCATACATTAATACTGGAAGACGTGTTTGTGGGGCTCTGCGCT 400
DB 241 CAACATCTCCAGCAGCAGTGCCTTTTCAACTGGAGGACGTCTATTGGGGCTATGTCT 300
QY 401 CGAAGGCTGAACATCAGATTCGAGGAGCTCCACTCCAGCCGACCTTTTTTCCAGGGGG 460
DB 301 TGACATATTGGACATTCACCTGGAGGAACCTCATACAGAGCAGACATTTCTTTCAGAGAG 360
QY 461 CTTACGCTTCTCGGTATGCTCTTCAGGAGGATCGTGGCTGCCACTTCATCAGGCTCG 520
DB 361 GCAGTCATTTCTCCGTTTGCAATAACAGTAACTTGTACGTCCCATGGAGTCAACACATA 420
QY 521 GACTCTCTGGACTACTGGCA 541
DB 421 TGAGAACATTTGATACTGGAA 441

RESULT 8
AJ507051 493 bp mRNA linear EST 17-SEP-2002
LOCUS AJ507051 Mus musculus BALB/c (Kahlem P) Mus musculus cDNA clone
DEFINITION AJ507051 182, mRNA sequence.
ACCESSION AJ507051
VERSION AJ507051.1 GI:23034136
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 493)
AUTHORS Gitton,Y., Dahmane,N., Baik,S., Ruiz i Altaba,A., Neidhardt,L., Scholze,M., Hermann,B.G., Kahlem,P., Ben Kahla,A., Schrinner,S., Yildirimman,R., Herwig,R., Lehrach,H. and Yaspo,M.L.
TITLE A gene expression map of human chromosome 21 orthologs in the mouse
JOURNAL Nature (2002) In press
COMMENT Contact: Kahlem P
Vertebrate Genomics
Max-Planck Institute for Molec. Genet.
Innestrass 73, D-14195 Berlin, GERMANY.
Location/Qualifiers
  1..493
  /organism="Mus musculus"

FEATURES
  source
    Location/Qualifiers
      1..493
        /organism="Mus musculus"
        /mol_type="mRNA"
        /db_xref="taxon:8364"
        /clone="THDA01ln08"
        /dev_stage="tailbud head (stage 28-30)"
        /lab_host="Escherichia coli DH10B"
        /clone_lib="XGC-tailbud-head"
        /note="Vector: pCSI07; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dT primed from sug of poly A+ RNA from tailbud head. EcoRI-NotI cut cDNA was then ligated into pCSI07 with EcoRI at the 5' end and NotI at the 3' end."
```



```
ORIGIN
  Query Match      25.4%; Score 148.2; DB 5; Length 626;
  Best Local Similarity 56.0%; Pred. No. 3.1e-33;
  Matches 304; Conservative 0; Mismatches 233; Indels 6; Gaps 1;

QY 3 GACATTATCCAGAAGGATTTCTAGACGCTTATTACAATCTGACCCCTGAAGACCATGATG 62
Db 83 GACATCATCCAGCAGGACTATCAGGACACATCTACAACCTCACCATCAAAACCCCTGATG 142
QY 63 GCATAGATGGTCCATCGCTTTGTCTCAGCGCGCTTTGTGATGAAACAGACTCA 122
Db 143 GGCATGAATCGGTGGCCACCATTGCCCCGCGGCGCACTATGTCTATGAAGACAGACAGT 202
QY 123 GACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACACC 182
Db 203 GACATGTTTGTAAACAGGAGTATCTCATCCAGAGCTCTGAGCCAGAGCTCCACCC 262
QY 183 -----AGTTTTTCATCTGGCTTTTGAACCTCAATGAGTTTCCCATCAGGAGCCATTTC 236
Db 263 AGACAGAGCTACTTCACTGGCTATCTGATGAGGGGTACGCCACCACCAACGCAACGAGGAC 322
QY 237 AGCAAGTGGTTTGTCAATGTAATCTGATATCCGTGGGACAGTACCCACCATCTCTGCTCC 296
Db 323 AGCAAGTGGTATCATGCTCCAGAGCTGTACCTTAGCGAGAGGTACCCCATCTTCTGCTCG 382
QY 297 GGCACCGGCTAGCTGTTTCTGGCGACGTGGCGAGTCAAGTGTACAACTGCTCCAAGAGC 356
Db 383 GGAACCGTTATGTATCTGGGACATCTGGCGAGAGTCTATCAGCGGTCTCTGAGC 442
QY 357 GTCCATACATTAACCTGAAGACGTTGTTGTGGGCTCTGCTCGAAGGCTGAACATC 416
Db 443 ATACGAGGCTGCACCTGAGGAGGTTTACGTGGGCATATGCTGGCCAAAGCTCCGCAAT 502
QY 417 AGATTGGAGGAGCTCACTCCAGCGGACCTTTTTCAGGGGGCTTAGCTTCTCCGTA 476
Db 503 GACCCCAACCGCCCGCCCAACAGTTCTCTTCAACCACTGAGGAGGTGTCTACTCCAGC 562
QY 477 TGCCCTTTCAGGAGGATCGTGGCTGCGCACTTCATCAAGCCTCGGACTCTCTTGGACTAC 536
Db 563 TGTAGTACAGCCACTCATACCTCCACCAAGTTCCAGCCCAACGAACTGGTCAAGTAC 622
QY 537 TGG 539
Db 623 TGG 625

RESULT 11
BJ740020/c
LOCUS BJ740020 MF015DA cDNA Oryzias latipes cdna clone MF015DA012002 3',
DEFINITION mRNA sequence.
ACCESSION BJ740020
VERSION BJ740020.1 GI:45337290
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 901)
AUTHORS Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished (2001)
COMMENT Contact: Tadao Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
  source
  Location/Qualifiers
    1..901
    /organism="Oryzias latipes"
    /mol_type="mRNA"
    /strains="Hd-rR"
    /db_xref="taxon:8090"
    /clone="MF015DA012002"
    /sex="mixture of female and male"
    /tissue_type="whole embryo"
    /dev_stage="organogenesis stage 35"
    /clone_lib="MF015DA cdna"

ORIGIN
  Query Match      22.4%; Score 131; DB 4; Length 901;
  Best Local Similarity 53.7%; Pred. No. 5.4e-28;
  Matches 297; Conservative 0; Mismatches 250; Indels 6; Gaps 1;

QY 1 GGGACATATTCAGAGGATTTCTTAGAGCTTATTAACAATCTGACCCCTGAAGACCATGA 60
Db 675 GTGACATCATCCAACAGGACTTTTGGATTTCTTATAAAAAATCTTATATAAAGACTGA 616
QY 61 TGGGCATAGATGGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAACAGACT 120
Db 615 TGGGATGAACTGGGTGGCAGTGCACCTGCAACGGGCAAGTTACGTCATGAAACAGACA 556
QY 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAA-----AACA 174
Db 555 GTGACATGTTTCTGCAACAGAGTACCTTATTTATAAGCTGCTGAAGCCAGAACTGAAG 496
QY 175 GAAACAAACAGGTTTTCATGCTGCTTCTGAAACTCAATGAGTTTCCCATCAGGAGCCCAT 234
Db 495 CCAAAACCAACTACTTCAACAGGAAACAACATGAGAGGCTTTTGACCCCAACCGGAATAAA 436
QY 235 TCAGCAAGTGGTTTCTGAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATCTGCT 294
Db 435 ACAGCAAGTGGTCAATGATCCCCCGAGCTGTACCCGAGGAAAGTACCCCATCTTTTGGCT 376
QY 295 CCGGCACCGGCTACGTGTTTCTGCGAGCTGCGGAGTCAGGTGAGTACAAATGTTCTCCAAGA 354
Db 375 CTGGAACCTGGCTATGTTCTCCGAGAGATGCGCGGAAATCTTACCAGACTTCTCTTGA 316
QY 355 GGTGCTCCATACATTAACCTGGAAGACGTTGTTGTGGGCTCTGCTCGAAGGCTGAACA 414
Db 315 GCACCCCGCCACCTGCACCTGGAGGACGTATAGCTGGGAATATGCTCTGGCCAAAGCTCAGAA 256
QY 415 TCAGATTGGAGGAGCTCCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTCTCCG 474
Db 255 TCGAGCCACCGCCCGCCCAATGGTTCCTGTTCAACCACTGGGCGGCTCTCTTACTCCA 196
QY 475 TATGCTCTTTCAGGAGGATCGTGGCTGCGACTTTCATCAAGCCTCGGACTCTCTTGGACT 534
Db 195 GCTGCAAGTACAGCCACCTGTAAATACATCAGTGGGTTCAGTCCCAATGAATACTGAAAT 136
QY 535 ACTGGAGGCTCT 547
Db 135 ACTGGCATCATCT 123

RESULT 12
AY401366
LOCUS AY401366
DEFINITION Mus musculus B3GALT1 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY401366
VERSION AY401366.1 GI:39757355
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 981)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
```



```

/misc_feature
/clone="9530002C17"
/sex="male"
/tissue_type="urinary bladder"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
284_1263
/notes="UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase,
polypeptide 1 (MGD) [GI:1349403, GB|NM_020283, evidence:
BLASTN, 99%, match=980]
putative"
polya_signal
2682..2687
/notes="putative"
polya_site
2705
/notes="putative"

```

# ORIGIN

```

Query Match      21.0%; Score 122.8; DB 3; Length 2705;
Best Local Similarity 59.3%; Pred. No. 2.3e-25;
Matches 249; Conservative 0; Mismatches 162; Indels 9; Gaps 2;

QY 3 GACATTATCCAGAAGGATTTCTAGACGCTATTACAAATCTGACCCCTGAAGACCATGATG 62
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 697 GACATCATCGTGGAGACTTCATTGACTCTACCACAATCTCACCTCTCAAACCTTAATG 756
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 63 GGCATAGATGGTCCATCGCTTTTCTCTCAGCGGCGCTTTGTGTGATGAAGAAACAGACTCA 122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 757 GGGATGAGATGGGTTGCCACTTCTCTTCAAAGCCAGTACGTATCATGAAGCCGACAGT 816
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 123 GACATGTTTCATCAGTTCGACTATCTGACTGAAGTCTCTGAA-----GAAACACAGA 176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 817 GACATTTTGTGAACATGACAACTTATTATTAACCTCTGAAACCCCTCTACCAAGCCA 876
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 177 ACAACGAGGTTTTTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGAGCCATTC 236
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 877 AGNAGAGATTTTCACTGGTTAGCTCA---TCAAGCGGGGCCAATCAGGATGTCCG 933
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 237 AGCAAGTGGTTGTGAGTAAATCTGAAATATCGGTGGGACAGTACCCACCATCTGCTCC 296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 934 AGTAAGTGGTATATGCTAGAGATTTGTACTCCCTGCAGCAACTACCCACCGTTCTGTTCA 993
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 297 GGCACCGGTCTAGTGTCTTCTGGCAGGTGGCGAGTCAAGTGTCAATGTCTCCAAGAGC 356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 994 GGGACTGGCTATATCTTTTCCGCTGATGTGGTGAACCTCAITTTACAAGACCTCGCTCCAC 1053
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 357 GTTCCCATACATTAACCTGAAGACGTTGTTGTGGGCTCTGCTCGAAAGGCTGAACATC 416
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1054 ACCAGGCTCTTCACTTGAAGATGTGTAGTGGGACTGTCTTCCGAAGCTTGGCATC 1113
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 14
BU133295
LOCUS      685 bp mRNA linear EST 25-NOV-2002
DEFINITION 603122104F1 CSEQCHL22 Gallus gallus cDNA clone CHEST85h2 5', mRNA
sequence.
ACCESSION  BU133295
VERSION    BU133295.1 GI:25345664
KEYWORDS   EST.
SOURCE     Gallus gallus (chicken)
ORGANISM   Gallus gallus

```

```

REFERENCE
AUTHORS    Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE      A Comprehensive Collection of Chicken cDNAs
JOURNAL    Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE    22335534
PUBMED     12445392
COMMENT    Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology
            (UMIST)

```

## FEATURES

source

```

1..585
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hixsex"
/db_xref="taxon:9031"
/clone="CHEST85h2"
/dev_stage="36"
/lab_host="DH10B"
/clone_lib="CSEQCHL22"
/notes="Organ: heads; Vector: pBluescript II KS(+); Site 1:
EcoRI; Site 2: NotI; Modification of pBluescript II KS(+):
[Stratagene] vector to accommodate cDNA produced with the
T-trimmed protocol (Construction of uni-directionally
cloned cDNA libraries from messenger RNA for improved 3'
end DNA sequencing by Glenn Fu, et al. U.S. Patent
6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI.
Ligate in double stranded adaptor containing BspI and
BamHI sites [5'ggcggcgtagcccgccggaacgaagaaag]
15'aattcttttttcggatccggggctgcacgc]"

```

## ORIGIN

```

Query Match      20.8%; Score 121.2; DB 5; Length 685;
Best Local Similarity 59.0%; Pred. No. 4.6e-25;
Matches 248; Conservative 0; Mismatches 163; Indels 9; Gaps 2;

QY 3 GACATTATCCAGAAGGATTTCTAGACGCTATTACAAATCTGACCCCTGAAGACCATGATG 62
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123 GACATTTATGTAGAGATTTTATTGACTCTCTATCACTCTCCTGAAACATTAAATG 182
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 63 GGCATAGATGGTCCATCGCTTTTGTCTCAGCGGCGCTTTGTGTGATGAAGAAACAGACTCA 122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183 GGGATGAGGTGGGTAGCAACATATTGTTTCAAAGCGGAAGTACGTTATGAAGACAGACAGT 242
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 123 GACATGTTTCATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAACAA-----GA 176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 243 GATATTTAGTAATAATATGGATAATCTTATTATAGTCTCTCAACCTTAACCAAGCCA 302
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 177 ACAACGAGGTTTTTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGAGCCATTC 236
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 303 AGGAGAAAGTACTTCACTGGTTACGTTA---TCAAGGAGGGCCAAATACGAGACGTTCCG 359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 237 AGCAAGTGGTTGTGAGTAAATCTGAAATATCGTGGGACAGTACCCACCATCTCTCTCC 296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 360 AGCAAGTGGTACATGCCAGAGATTTGTATCTCTGACAGCAANTTACCCACCTTCTGTTCCG 419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 297 GGCACCGGTCTAGTGTCTTCTGGCAGCTGGCGAGTCAAGTGTACAAATGTCTCCAAGAGC 356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 420 GGCATCGGCTACATTTTTCAGCAGATGTAGCAGAACTGATTTACAAACCTCCCTTCCAC 479
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 357 GTTCCCATACATTAACCTGAAGACGTTGTTGTGGGCTCTGCTCGAAAGGCTGAACATC 416
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 480 ACCAGACTCTCTCACTCTGAAGACGTTGATGTGCGGACTCTGTCTCGGAAGCTGGGCAATC 539
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 15
BU350312
LOCUS      842 bp mRNA linear EST 28-NOV-2002
DEFINITION 603527845F1 CSEQCHN69 Gallus gallus cDNA clone CHEST47712 5', mRNA
sequence.
ACCESSION  BU350312
VERSION    BU350312.1 GI:25858313
KEYWORDS   EST.
SOURCE     Gallus gallus (chicken)
ORGANISM   Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE  1 (bases 1 to 842)

```

AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT

Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
Pong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
A Comprehensive Collection of Chicken cDNAs  
Curr. Biol. 12 (22), 1965-1969 (2002)

22335534  
12445392

Contact: Simon Hubbard

Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)

PO Box 88, Manchester, M60 1QD, UK

Tel: 01612008930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

Location/Qualifiers

1..842

/organism="Gallus gallus"

/mol\_type="mRNA"

/strains="Compton Line 151"

/db\_xref="taxon:9031"

/clone="ChEST477112"

/sex="Female"

/tissue type="cerebellum"

/dev stage="adult"

/lab\_host="DH10B"

/clone\_lib="CSFQCHN69"

/note="Organ: brain; Vector: pBluescript II KS(+); Site 1:  
EcoRI; Site 2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was blunted, ligated to NotI adapters, digested with  
EcoRI, size-selected, and cloned into the NotI and EcoRI  
compatible site of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
(1996): 791, except that a significantly longer  
reannealing hybridization was used."

FEATURES  
source

## ORIGIN

Query Match 20.8%; Score 121.2; DB 5; Length 842;  
Best Local Similarity 59.0%; Pred. No. 4.9e-25;  
Matches 248; Conservative 0; Mismatches 163; Indels 9; Gaps 2;  
QY 3 GACATTATCCAGAAGATTCTCTAGACGCTATTACAAATCTGACCCCTGAAGACCATGATG 62  
Db 21 GACATTATTTGTAGAGGATTTTATTTGACTCTCTATCATACCTCACTCTGAACACATTATG 80  
QY 63 GGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGGCGGCTTTGTGATGAAACACACTCA 122  
Db 81 GGGATGAGGTGGGTAGCAATATTTCTCAAAGCGAAGTAGCTTTATGAAGACAGACAGT 140  
QY 123 GACATGTTTCATCAATGTTTCACTATCTGACTGCTCTCTGAAGAAACA-----GA 176  
Db 141 GATAATTTTGTAAATATGATATCTTATTTAAGCTGCTCAACCTAACACCAAGCCA 200  
QY 177 ACAACCAAGCTTTTTCACCTGGCTTTCTTGAACCTCAATGAGTTTCCCATCAGGCAGCCATT 236  
Db 201 AGGAGAAGGTACTTCACTGTTACGTTA---TCAACGGAGGGCCAATACGAGACGTTGCG 257  
QY 237 AGCAAGTGGTTTGTCTAGTAAATCTGAAATATCCGTGGGACAGGTACCCACCATTTCTGCTCC 296  
Db 258 AGCAAGTGGTACATGCCAGAGATTGTATCTCTGACAGCAATTTACCCACCTTCTGTTGCG 317  
QY 297 GGCACGGCTAGCTGTTTCTGCGACGTGGCGAGTCAAGTGTACATGCTCTCCAAGAGC 356  
Db 318 GGCACCTGGCTACATTTTTCAGCAGATGTAGCAGAACTGATTTTACAAAACCTCCCTTCAC 377  
QY 357 GTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCGCTCGAAAGGCTGAACATC 416  
Db 378 ACCAGACTCTCCATCTCGAAGACGTGTATGTCGGACTCTGTCTTCGGAAGCTGGGCATC 437

Search completed: September 22, 2005, 16:56:10  
Job time : 2583.06 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2005, 12:15:17 ; Search time 121.581 Seconds  
(without alignments)  
7859.644 Million cell updates/sec

Title: US-10-777-828-8\_COPY\_428\_1011

Perfect score: 584

Sequence: 1 gggacattatccagaaggat.....aagattgtccgctgtctga 584

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PTCUS COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	584	100.0	1011	US-09-831-630-8	Sequence 8, Appli
2	80.8	13.8	1134	US-09-482-180A-3	Sequence 3, Appli
3	56.2	9.6	1191	US-09-459-133-3	Sequence 3, Appli
4	53	9.1	1167	US-09-459-133-14	Sequence 14, Appli
5	52.4	9.0	1464	US-09-774-528-430	Sequence 430, Appli
6	48	8.2	1420	US-09-482-180A-1	Sequence 1, Appli
7	48	8.2	22374	US-09-949-016-14938	Sequence 14938, A
8	46	7.9	125902	US-09-949-016-13715	Sequence 13715, A
9	40.2	6.9	1434	US-09-055-097-2	Sequence 2, Appli
10	40.2	6.9	1434	US-09-373-902-2	Sequence 2, Appli
11	40.2	6.9	15640	US-09-949-016-5696	Sequence 5696, Ap
12	40.2	6.9	15640	US-09-949-016-17438	Sequence 17438, A
13	40.2	6.9	8436	US-09-949-016-13963	Sequence 13963, A
14	39.2	6.7	748	US-09-270-767-444	Sequence 444, App
15	39.2	6.7	748	US-09-270-767-15726	Sequence 15726, A
16	38.6	6.6	17731	US-09-949-016-16365	Sequence 16365, A
17	37	6.3	2359	US-09-620-312D-699	Sequence 699, App
18	37	6.3	2426	US-09-620-312D-697	Sequence 697, App
19	36.6	6.3	1098	US-09-716-793A-3	Sequence 3, Appli
20	35.8	6.1	374	US-09-270-767-1385	Sequence 1385, Ap
21	35.8	6.1	374	US-09-270-767-16667	Sequence 16667, A
22	35.4	6.1	1279	US-09-902-540-6180	Sequence 6180, Ap
23	35.4	6.1	1959	US-09-902-540-313	Sequence 313, App
24	35	6.0	2557	US-08-464-954A-1	Sequence 1, Appli
25	34.4	5.9	861	US-09-347-613C-3	Sequence 3, Appli
26	34.4	5.9	861	US-09-662-183A-3	Sequence 3, Appli
27	34.2	5.9	1100	US-09-248-335-53	Sequence 53, Appli

ALIGNMENTS

RESULT 1

US-09-831-630-8  
; Sequence 8, Application US/09831630  
; Patent No. 6800468  
; GENERAL INFORMATION:  
; APPLICANT: Clausen, Henrik  
; APPLICANT: Amado, Margarita  
; TITLE OF INVENTION: UDP-GALACTOSE: BETA-N-ACETYL-GLUCOSAMINE BETA 1,3  
; FILE REFERENCE: 7188-157  
; CURRENT APPLICATION NUMBER: US/09/831,630  
; CURRENT FILING DATE: 2001-07-12  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 1011  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (79)..(1008)  
US-09-831-630-8

Query Match	100.0%	Score 584;	DB 4;	Length 1011;
Best Local Similarity	100.0%;	Pred. No. 1.2e-183;		
Matches 584;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	1	GGGACATTATCCAGAAAGGATTTCTTAGACGCTATTACAACTGACCTCGAAGACCATGA	60	
Db	428	GGGACATTATCCAGAAAGGATTTCTTAGACGCTATTACAACTGACCTCGAAGACCATGA	487	
Qy	61	TGGGCATAGAAATGGGTCCATCGCTTTTCTCCTCAGCGCGGTTTGTGTGATGAAACAGACT	120	
Db	488	TGGGCATAGAAATGGGTCCATCGCTTTTCTCCTCAGCGCGGTTTGTGTGATGAAACAGACT	547	
Qy	121	CAGACATGTTTCATCAATGTTGACTATCTGACTGAACCTCTTCTGAAGAAAACAGAACAA	180	
Db	548	CAGACATGTTTCATCAATGTTGACTATCTGACTGAACCTCTTCTGAAGAAAACAGAACAA	607	
Qy	181	CCAGGTTTTTCACCTGGCTTCTTGAACCACTCAATGAGTTTCCCATCAGGAGCCATTCAGCA	240	
Db	608	CCAGGTTTTTCACCTGGCTTCTTGAACCACTCAATGAGTTTCCCATCAGGAGCCATTCAGCA	667	
Qy	241	AGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGTACCCACATTTGCTCGGCA	300	
Db	668	AGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGTACCCACATTTGCTCGGCA	727	
Qy	301	CCGGTAGCTGTTTTCGGCCACGCTGGCGAGTCAGGTGTACATGTTCTCCAGAGCGTCC	360	
Db	728	CCGGTAGCTGTTTTCGGCCACGCTGGCGAGTCAGGTGTACATGTTCTCCAGAGCGTCC	787	

QY 361 CATACATTAAGTGAAGACGTTGTTGGGGCTCTGCTCGAAGGCTGAACATCAGAT 420  
Db CATACATTAAGTGAAGACGTTGTTGGGGCTCTGCTCGAAGGCTGAACATCAGAT 847  
QY 421 TGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGGCTTCTCCGATGCCC 480  
Db TGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGGCTTCTCCGATGCCC 907  
QY 481 TCTTCAGAGAGATCGTGGCCCTGCCACTTTCATCAAGCCTCGGACTCTCTTGGACTACTGGC 540  
Db TCTTCAGAGAGATCGTGGCCCTGCCACTTTCATCAAGCCTCGGACTCTCTTGGACTACTGGC 967  
QY 541 AGGCTCTAGAGATTCGCGGGGGAAGATTCTCGCCCTGTCTGA 584  
Db AGGCTCTAGAGATTCGCGGGGGAAGATTCTCGCCCTGTCTGA 1011  
RESULT 2  
US-09-482-180A-3  
; Sequence 3, Application US/09482180A  
; Patent No. 6361985  
; GENERAL INFORMATION:  
; APPLICANT: Konklin, Darrell C.  
; APPLICANT: Yamamoto, Gayle  
; APPLICANT: Gao, Zeren  
; APPLICANT: Whitmore, Theodore E.  
; APPLICANT: Jaepers, Stephen  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING  
; TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOG, ZN5SP6  
; FILE REFERENCE: 98-80  
; CURRENT APPLICATION NUMBER: US/09/482,180A  
; CURRENT FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/115,721  
; PRIOR FILING DATE: 1999-01-12  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 1134  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: degenerate sequence  
; NAME/KEY: variation  
; LOCATION: (1)...(1134)  
; OTHER INFORMATION: n is any nucleotide  
; NAME/KEY: misc feature  
; LOCATION: (1)...(1134)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-482-180A-3  
Query Match 13.8%; Score 80.8; DB 3; Length 1134;  
Best Local Similarity 32.2%; Pred. No. 3.8e-16;  
Matches 134; Conservative 61; Mismatches 221; Indels 0; Gaps 0;  
QY 3 GACATATCCAGAGGATTCCTAGACGCTCTATACATCTGACCCCTGAAGACCATGATG 62  
Db GAYATHYNTCAATGGGAYTTACNGARGAYTTTAAAYTNACNYTNAARGAYTNCAY 597  
QY 63 GGCATAGAATGGTCCATCGCTTTTGTCTCAGGGGGGTTTGTGATGAAACACAGATCA 122  
Db YTNCAKMGNTGGGTGNGCNGTGYCCNCAAGNCAYTYTAYGTYNARGGNGAYGAY 657  
QY 143 GACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTTGAAGAAAAACAGAACACC 182  
Db GAYGNTTYGTNCAYGTGTCNAAAYGTNTGARTTYTNGAYGNTGGAYCCNGCNCAR 717  
QY 143 AGGTTTTTCACTGGCTTCTTGAACCTCAATGATTTCCCATCAGCAGCCATTGAGCAAG 242  
Db GAYYNTYNTGNGNGAYGTNATHMGNCAAGCNGNYTNCNAAAYMGNAAYACNAAAGTNAAR 777  
QY 243 TGGTTTGTCAAGTAAATCTGAATATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCACC 302

Db 778 TAVTATVTHCCNCNWSNATGTATYMGNGCNACNCAVATYCCNCCNTAYCGNGGNGNGCN 837  
QY 303 GGCTACGCTGTTTCTGCGGACGTGGCGAGTCAGGTGTACAAATGTCTCCAAGAGCGTCCCA 362  
Db 838 GENTAYGTNATGWSNMGNGCNACNGTNNMGNGNYTNCARGCNATHATGGARGAYGCNGAR 897  
QY 363 TACATTAACCTGGAAGACGCTGTTTGTGGGGCTCTGCCCTCGAAAGGCTGAACATCAG 418  
Db 898 YTYTWSNATHGAYGAYGNTTYGTNGGNATGTGYYTNNMGNGNYTNGGNYTWS 953  
RESULT 3  
US-09-459-133-3  
; Sequence 3, Application US/09459133  
; Patent No. 8416988  
; GENERAL INFORMATION:  
; APPLICANT: Konklin, Darrell C.  
; APPLICANT: Yamamoto, Gayle  
; APPLICANT: Jaepers, Stephen R.  
; APPLICANT: Gao, Zeren  
; TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS  
; FILE REFERENCE: 98-77  
; CURRENT APPLICATION NUMBER: US/09/459,133  
; CURRENT FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: 60/111,697  
; PRIOR FILING DATE: 1998-12-10  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 1191  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Degenerate sequence  
; NAME/KEY: misc feature  
; LOCATION: (1)...(1191)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-459-133-3  
Query Match 9.6%; Score 56.2; DB 3; Length 1191;  
Best Local Similarity 28.5%; Pred. No. 6e-08;  
Matches 161; Conservative 73; Mismatches 322; Indels 9; Gaps 2;  
QY 1 GGGACATATCCAGAGATTCCTAGACGCTCTATACATCTGACCCCTGAAGACCATGA 60  
Db 620 SNGAYTYNTYNTGTTGGGAYTTTGTNGAYTNCCTTTTAAAYCARACNYTNAARGAYTNY 679  
QY 61 TGGGCATAGAATGGTCCATCGCTTTTGTCTCAGGGGGGTTTGTGATGAAACACAGACT 120  
Db 680 TNYTNYTNGCNGTGGYTNMGNGNCAYTYGNCNACNGTNSNTTYGTNYTNGCNGCARG 739  
QY 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTCTGAAACATCAATGAGTTTCCCATCAGGAGCCAT 174  
Db 740 AYGAYGNTTYGTNCAYACNCCNGCNYTNTYNGCNCAYTYTNGMNGCNYTNCNCCNGCNW 799  
QY 175 GAACACACAGGTTTTTCACTGGGCTTCTGAAACATCAATGAGTTTCCCATCAGGAGCCAT 234  
Db 800 SNGCMNGWSNYTNTAYTYTNGGNGARGTNTTYACNCAAGCNGATGCCNYTNNMNAARCCNG 859  
QY 235 TCAGCAAGTGGTTGTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACATCTCTGCT 294  
Db 860 GNGNCCNTTYATGTCNCGARGNSNTTYTYTGARGNG---GNTAYCCNCCNTAYGNCW 916  
QY 295 CCGCACCGGCTCAGTGTGTTTCTGGCGACGTCGGCAGTCAGGTGTACAATGTCTTCCAAGA 354  
Db 917 SNGGNGGNGNTAYGTNATHGNGNGMNGYTNCCNCCNTGGYTYTNNMGNGCNGCNGNM 976  
QY 355 GCGTCCCATACATTAATAACTGGAACACGCTGTTGTGGGCTCTGCTCGAAAGGCTGAACA 414  
Db 977 GNGTNGCNCNTTYCCNTTYTGARGAYGTNTAYACNGGNYTNTGYATHMNGCNGTNGNY 1036  
QY 415 TCAGATTGGAGAGCTCCACTCCCGACGACCTTTTTTCCAGGGGGGCTTACGCTTCTCCG 474

Db 1037 TNGTNCNCARCNCAAYCNGGNTTYYTNACNGCNTGGCCNCGNAYMGNAACNGCNGAYC 1096  
Qy 475 TATGCTCTTTCAGGAGGATCGTGGCTGCACCTTCAACGCTCGGACTCTCTTGGACT 534  
Db 1097 AVTGYCNTTYMGNAAYTNTYNTGNTMGNCNTYNGCCNCCARCGNWSNATHMGNY 1156  
Qy 535 ACTGGCAGGCTCTAGAGAAATCCCG 559  
Db 1157 TNTGGAARCAARYTNCARGAYCCNMG 1181

RESULT 4  
US-09-459-133-14  
; Sequence 14, Application US/09459133  
; Patent No. 641698  
; GENERAL INFORMATION:  
; APPLICANT: Conklin, Darrell C.  
; APPLICANT: Yamamoto, Gayle  
; APPLICANT: Jaspers, Stephen R.  
; APPLICANT: Gao, Zeren  
; TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS  
; FILE REFERENCE: 98-77  
; CURRENT APPLICATION NUMBER: US/09/459,133  
; CURRENT FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: 60/111,697  
; PRIOR FILING DATE: 1998-12-10  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 14  
; LENGTH: 1167  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: degenerate sequence  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(1167)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-459-133-14

Query Match 9.1%; Score 53; DB 3; Length 1167;  
Best Local Similarity 32.4%; Pred. No. 6.9e-07;  
Matches 132; Conservative 53; Mismatches 213; Indels 9; Gaps 2;  
Qy 1 GGGCATTATCCAGAAGGATTTCTAGACGCTATTACAATCTGACCTCGAAGACCATGA 60  
Db 596 GNGAYTNTYNTGCGAYTYYTNGAYTNCCTATAYATMGNAACNTYNAARGAYTNY 655  
Qy 61 TGGGCATAGATGGTCCATCGCTTTTCTCTCAGCGCGCGTTTGTGATGAAACAGACT 120  
Db 656 TNYTNTACNTGGYTNSNCAYCAYTCYCCNGAYGTNAAYTYYTNTYNTNARGTNCARG 715  
Qy 121 CAGACATGTTCAATCAATGTGACTATCTGACTGAATCTGCTCTGAGAA-----AAACA 174  
Db 716 AYGAYGCTTGTNCAYATHCNCGNCTYNTYNGARCAYYTNCARACNTYNCNCNCNAGT 775  
Qy 175 GAACAACAGGTTTTCTACTGCTTCTTGAACCTCAATGAGTTTCCCATCAGGACCAT 234  
Db 776 GGGCNMGNSYNTAYTNGNGGARATHTTYTACNARGCNAARCCNTYNTMGNAARCCNG 835  
Qy 235 TCAGCAAGTGGTTTGTAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATCTGCT 294  
Db 836 GNGNCCNTTYTAYTNCNARACNTYYTYGARGNGAY---TAYCNGNTAYGCNW 892  
Qy 295 CCGGCACCGGTACGTGTTTTCTGCGGACGTGGGAGTCAGGTGACAAATGTCTCCAAGA 354  
Db 893 SNGGNGGNGNTAYGTNATHWSNMGNTYNGCNCNTGGYNTYNTNARGCNGCNGCNM 952  
Qy 355 CGGTCCCATACATTAACCTGAGACGCTTTGTGGGCTCTGCCTC 401  
Db 953 GNGTNGCNCNTTYCCNTTYGAYGAYGTNTAYACNGGNTTYGTYY 999

RESULT 5  
US-09-774-528-430  
; Sequence 430, Application US/09774528  
; Patent No. 6743619  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Zhou, Ping  
; APPLICANT: Goodrich, Ryle  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Zhang, Jie  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6743619el Nucleic Acids and  
; FILE REFERENCE: 802  
; CURRENT APPLICATION NUMBER: US/09/774,528  
; CURRENT FILING DATE: 2001-01-30  
; NUMBER OF SEQ ID NOS: 441  
; SOFTWARE: pt\_FL\_genes Version 2.0  
; SEQ ID NO 430  
; LENGTH: 1464  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1464)  
US-09-774-528-430

Query Match 9.0%; Score 52.4; DB 4; Length 1464;  
Best Local Similarity 47.5%; Pred. No. 1.2e-06;  
Matches 189; Conservative 0; Mismatches 206; Indels 3; Gaps 1;  
Qy 3 GACATTATCCAGAAGGATTTCTAGACGCTATTACATCTGACCTGAAGACCATGATG 62  
Db 613 GATATAATTGAAGGAATCTTCTTGACAGAGTTCTGAGAACCAACCTGAAGATCATTGCA 672  
Qy 63 GGCATAGAAATGGGTCCATCGCTTTTGTCTCAGCGCGGTTTGTGATGAAACAGACTCA 122  
Db 673 ATGATACAGTGGGCTGTGGCTTTCTGCCCTATATGCCCTGTTTCATTCTCAAGGTGATGAA 732  
Qy 123 GACATGTTTCATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGAAAACAGAACACC 182  
Db 733 GAGAGCTTTGTCAATCTACCAAGCTTGGTAGACTATCTTCTCAATCTGAAGAACACCTA 792  
Qy 183 AGGTTTTTCATCGCTTCTTGAACCTCAATGAGTTT---CCCATCAGGACGACCATTCAGC 239  
Db 793 GAAGATATCTATGTAGGAAGAGTTTCTTTCATCAGGTTACACCCCAATAGAGATCTCTCAGAAC 852  
Qy 240 AAGTGGTTTGTTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATCTGCTCCGCG 299  
Db 853 AGAGACTTTTGCCTCTTAGTGAGTACCCAGAAAATACTACCCAGATTACTGAGTGTGT 912  
Qy 300 ACCGCTACGTGTTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTTCCAAGAGCGTC 359  
Db 913 GAGGCTTTATAATGTCCCAAGATGTGGCTCGAATGATGTATGTGTTTTTCAAGAAGATGA 972  
Qy 360 CCATACATTAACCTGGAAGACGTGTTTGTGGGCTCTG 397  
Db 973 CCCATGATGGTCCAGCTGATGTGTTTGTAGGAATTTG 1010

RESULT 6  
US-09-482-180A-1  
; Sequence 1, Application US/09482180A  
; Patent No. 6361985

```

; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Yamamoto, Gayle
; APPLICANT: Gao, Zeren
; APPLICANT: Whitmore, Stephen
; APPLICANT: Jaspers, Stephen
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING
; TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOG, ZNSSP6
; FILE REFERENCE: 98-80
; CURRENT APPLICATION NUMBER: US/09/482,180A
; CURRENT FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/115,721
; PRIOR FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1420
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (135)...(1271)
; US-09-482-180A-1

Query Match      8.2%; Score 48; DB 3; Length 1420;
Best Local Similarity 44.7%; Pred. No. 3.6e-05;
Matches 186; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

Qy 3 GACATATCCAGAGGATTTCTAGACGTCTATTACAACTCTGACCCCTGAAGACCATGATG 62
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 572 GACATCTCCAGTGGGACTTCACTGAGGACTTCTTCAACCTGACGCTCAAGGAGCTGCAC 731

Qy 63 GGCATAGAAATGGTCCATCGCTTTTGTCTCAGGGGGGTTTGTGATGAAGAAACAGACTCA 122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 732 CTGCAGCGCTGGGTGGTGGCTCGCTGCCGCCAGGCCCAATTTTCATGCTAAAGGAGATGAC 791

Qy 123 GACATGTTTCATCAATGTTGACTATCTGACTGAAGTCTTCTGAAGAAAAACAGAACCAACC 182
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 792 GATGCTTTGTCCAGTCCCAACGTTGATAGATTCCTGGATGGTGGGACCCAGCCAG 851

Qy 183 AGGTTTTTCACTGGCTTCTTGAAGAACTCAATGAGTTTCCCATCAGGACGCCATTTCAGCAAG 242
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 854 GACCTCTCGTGGGAGATGTATCGGCCAAGCGCTGCCCAACAGGAACACACTAAGGTCAAA 911

Qy 243 TGGTTTGTCAAGTAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCGGGCACC 302
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 912 TACTTTCATCCCAACCCCTCAATGTACAGGGCCACCCCACTACCCACCCCTATGCTGGTGGGGGA 971

Qy 303 GGCTACGTGTTTCTGGCGAGCTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCA 362
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 972 GGATATGTCATGTCAGAGCCACAGTGGCGGCTCCAGGCTATCATGGAAGATGCTGAA 1031

Qy 363 TACATTAACCTGGAAGACGTGTTTGTGGGGCTCTGCCCTCGAAAGGCTGAACATCAG 418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1032 CTCCTCTCCATTGATGATGCTCTTTGTGGGTATGTCCCTGAGGAGGCTGGGGCTGAG 1087

RESULT 7
US-09-949-016-14938/c
; Sequence 14938, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13715
; LENGTH: 125902
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(125902)
; OTHER INFORMATION: n = A,T,C or G

; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14938
; LENGTH: 22374
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(22374)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14938

Query Match      8.2%; Score 48; DB 4; Length 22374;
Best Local Similarity 44.7%; Pred. No. 0.00019;
Matches 186; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

Qy 3 GACATATCCAGAGGATTTCTAGACGTCTATTACAACTCTGACCCCTGAAGACCATGATG 62
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 21354 GACATCTCCAGTGGGACTTCACTGAGGACTTCTTCAACCTGACGCTCAAGGAGCTGCAC 21295

Qy 63 GGCATAGAAATGGTCCATCGCTTTTGTCTCAGGGGGGTTTGTGATGAAGAAACAGACTCA 122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 21294 CTGCAGCGCTGGGTGGTGGCTGCTGCCGCCAGGCCCAATTTTCATGCTAAAGGAGATGAC 21235

Qy 123 GACATGTTTCATCAATGTTGACTATCTGACTGAAGTCTTCTGAAGAAAAACAGAACCAACC 182
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 21234 GATGCTTTGTTCACGTCCCAACGTTGATAGATTCCTGGATGGTGGGACCCAGCCAG 21175

Qy 183 AGGTTTTTCACTGGCTTCTTGAAGAACTCAATGAGTTTCCCATCAGGACGCCATTTCAGCAAG 242
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 21174 GACCTCTCGTGGGAGATGTATCGGCCAAGCGCTGCCCAACAGGAACACACTAAGGTCAAA 21115

Qy 243 TGGTTTGTCAAGTAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCGGGCACC 302
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 21114 TACTTTCATCCCAACCCCTCAATGTACAGGGCCACCCCACTACCCACCCCTATGCTGGTGGGGGA 21055

Qy 303 GGCTACGTGTTTCTGGCGAGCTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCA 362
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 21054 GGATATGTCATGTCAGAGCCACAGTGGCGGCTCCAGGCTATCATGGAAGATGCTGAA 20995

Qy 363 TACATTAACCTGGAAGACGTGTTTGTGGGGCTCTGCCCTCGAAAGGCTGAACATCAG 418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20994 CTCITCCCAATTGATGATGCTCTTTGTGGGTATGTGCTGAGGAGGCTGGGGCTGAG 20939

RESULT 8
US-09-949-016-13715/c
; Sequence 13715, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13715
; LENGTH: 125902
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(125902)
; OTHER INFORMATION: n = A,T,C or G
```

US-09-949-016-13715

Query Match 7.9%; Score 46; DB 4; Length 125902;  
 Best Local Similarity 52.0%; Pred. No. 0.0026;  
 Matches 103; Conservative 0; Mismatches 95; Indels 0; Gaps 0;  
 QY 29 CGTCTATTACATCTGACCTGAAGACCATGATGGGCTAGATGGGTCATCGCTTTTG 88  
 Db 9845 CGTCTCAAAACAACAACAAAAACCAACCTATTATGGCATTCAGGTGGGTAACTGAGTTTG 9786  
 QY 89 TCCTCAGCGGGGTTGTGTGATGAAACAGACTCAGACATGTTTCATCAATGTTGACTATCT 148  
 Db 9785 CCCCATTGCCAATCATCATGACAGACACATTTATGTGTTTCATCAATGCGCAATTT 9726  
 QY 149 GACTGAACCTGCTTCTGAAGAAAAACAGAACACCGAGTTTTCATCTGGCTTCTTGAAACT 208  
 Db 9725 AGTGTAGTATCTTTTAAACTTAAACCACTCAGAGAAGTTTTTTCACAGGTTATCCCTAAT 9666  
 QY 209 CAATGAGTTTCCCATCAG 226  
 Db 9665 TGATAAATTATCCCATAG 9648

RESULT 9

US-09-055-097-2  
 ; Sequence 2, Application US/09055097  
 ; Patent No. 5955282  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hillman, Jennifer L.  
 ; APPLICANT: Guegler, Karl J.  
 ; APPLICANT: Corley, Neil C.  
 ; APPLICANT: Shah, Purvi  
 ; APPLICANT: Patterson, Chandra  
 ; TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSEQ for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/055,097  
 ; FILING DATE: Filed Herewith  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Cerrone, Michael C.  
 ; REGISTRATION NUMBER: 39,132  
 ; REFERENCE/DOCKET NUMBER: PF-0490 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (650) 855-0555  
 ; TELEFAX: (650) 845-4166  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1434 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: DUODNOT02  
 ; CLONE: 1705085

US-09-055-097-2 Query Match 6.9%; Score 40.2; DB 2; Length 1434;

Best Local Similarity 54.4%; Pred. No. 0.014;  
 Matches 81; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
 QY 1 GGGACATTATCCAGAAGGATTTCCTAGACGCTCTATTACAATCTGACCTCGAAGACCATGA 60  
 Db 446 GGGATATCTTCAGCGCGCTTCCAGGACTCTTACCGCACTCACCCTTAAGACCTCA 505  
 QY 61 TGGGCATAGAAATGGTCCATCGCTTTTGTCTCAGGCGCGTTTGTGATGAAAAACAGACT 120  
 Db 506 GCGGCTGNACTGGGCTGAGAAACACTGCCCATGCGCCGATACGTCCTCAAGACGGACG 565  
 QY 121 CAGACATGTTTCATCAATGTTGACTATCTG 149  
 Db 566 ATGATGTGTATGTCAACGTCCTCGAATG 594

RESULT 10

US-09-373-902-2  
 ; Sequence 2, Application US/09373902  
 ; Patent No. 6649737  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hillman, Jennifer L.  
 ; APPLICANT: Guegler, Karl J.  
 ; APPLICANT: Corley, Neil C.  
 ; APPLICANT: Shah, Purvi  
 ; APPLICANT: Patterson, Chandra  
 ; TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSEQ for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/373,902  
 ; FILING DATE: 12-Aug-1999  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/055,097  
 ; FILING DATE: <Unknown>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Cerrone, Michael C.  
 ; REGISTRATION NUMBER: 39,132  
 ; REFERENCE/DOCKET NUMBER: PF-0490 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (650) 855-0555  
 ; TELEFAX: (650) 845-4166  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1434 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: DUODNOT02  
 ; CLONE: 1705085  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-373-902-2 Query Match 6.9%; Score 40.2; DB 4; Length 1434;  
 Best Local Similarity 54.4%; Pred. No. 0.014;  
 Matches 81; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
 QY 1 GGGACATTATCCAGAAGGATTTCCTAGACGCTCTATTACAATCTGACCTCGAAGACCATGA 60  
 Db 446 GGGATATCTTCAGCGCGCTTCCAGGACTCTTACCGCACTCACCCTTAAGACCTCA 505

```
Qy 61 TGGCATAGATGGTCCATCGCTTTTGTCTCTCAGCGCGCTTTTGTGATGAACAGACT 120
Db 506 GCGGCTGAATGGGCTGAGAAACACTGCCCATGGCCCGATACGTCTCTCAAGACGGACG 565
Qy 121 CAGACATGTTTCATCAATGTTGACTATCTG 149
Db 566 ATGATGTGTATGTCAACGCTCCCTGACTG 594

RESULT 11
US-09-949-016-5696
; Sequence 5696, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5696
; LENGTH: 1640
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5696

Query Match 6.9%; Score 40.2; DB 4; Length 1640;
Best Local Similarity 54.4%; Pred. No. 0.015; Indels 0; Gaps 0;
Matches 81; Conservative 0; Mismatches 68;

Qy 1 GGGACATATCCAGAAGGATTTCCTAGACGCTTATTACAATCTGACCGCTGAAGACCATGA 60
Db 838 GGGATATCTTGCAGGCGCCCTTCCAGGACTCTACCGCAACTCAGCTCAACCTAAAGACCCCTCA 697
Qy 61 TGGCATAGATGGTCCATCGCTTTTGTCTCTCAGCGCGCTTTTGTGATGAACAGACT 120
Db 598 GCGGCTGAATGGGCTGAGAAACACTGCCCATGGCCCGATACGTCTCTCAAGACGGACG 757
Qy 121 CAGACATGTTTCATCAATGTTGACTATCTG 149
Db 758 ATGATGTGTATGTCAACGCTCCCTGACTG 786

RESULT 12
US-09-949-016-17438
; Sequence 17438, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17438
; LENGTH: 5640
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17438

Query Match 6.9%; Score 40.2; DB 4; Length 8436;
Best Local Similarity 54.4%; Pred. No. 0.042; Indels 0; Gaps 0;
Matches 81; Conservative 0; Mismatches 68;

Qy 1 GGGACATATCCAGAAGGATTTCCTAGACGCTTATTACAATCTGACCGCTGAAGACCATGA 60
Db 7758 GGGATATCTTGCAGGCGCCCTTCCAGGACTCTACCGCAACTCAGCTCAACCTAAAGACCCCTCA 7817
Qy 61 TGGCATAGATGGTCCATCGCTTTTGTCTCTCAGCGCGCTTTTGTGATGAACAGACT 120
Db 7818 GCGGCTGAATGGGCTGAGAAACACTGCCCATGGCCCGATACGTCTCTCAAGACGGACG 7877
Qy 121 CAGACATGTTTCATCAATGTTGACTATCTG 149
Db 7878 ATGATGTGTATGTCAACGCTCCCTGACTG 7906

RESULT 14
US-09-270-767-444
; Sequence 444, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
```

```
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17438

Query Match 6.9%; Score 40.2; DB 4; Length 5640;
Best Local Similarity 54.4%; Pred. No. 0.033; Indels 0; Gaps 0;
Matches 81; Conservative 0; Mismatches 68;

Qy 1 GGGACATATCCAGAAGGATTTCCTAGACGCTTATTACAATCTGACCGCTGAAGACCATGA 60
Db 2638 GGGATATCTTGCAGGCGCCCTTCCAGGACTCTACCGCAACTCAGCTCAACCTAAAGACCCCTCA 2697
Qy 61 TGGCATAGATGGTCCATCGCTTTTGTCTCTCAGCGCGCTTTTGTGATGAACAGACT 120
Db 2698 GCGGCTGAATGGGCTGAGAAACACTGCCCATGGCCCGATACGTCTCTCAAGACGGACG 2757
Qy 121 CAGACATGTTTCATCAATGTTGACTATCTG 149
Db 2758 ATGATGTGTATGTCAACGCTCCCTGACTG 2786

RESULT 13
US-09-949-016-13963
; Sequence 13963, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13963
; LENGTH: 8436
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13963

Query Match 6.9%; Score 40.2; DB 4; Length 8436;
Best Local Similarity 54.4%; Pred. No. 0.042; Indels 0; Gaps 0;
Matches 81; Conservative 0; Mismatches 68;

Qy 1 GGGACATATCCAGAAGGATTTCCTAGACGCTTATTACAATCTGACCGCTGAAGACCATGA 60
Db 7758 GGGATATCTTGCAGGCGCCCTTCCAGGACTCTACCGCAACTCAGCTCAACCTAAAGACCCCTCA 7817
Qy 61 TGGCATAGATGGTCCATCGCTTTTGTCTCTCAGCGCGCTTTTGTGATGAACAGACT 120
Db 7818 GCGGCTGAATGGGCTGAGAAACACTGCCCATGGCCCGATACGTCTCTCAAGACGGACG 7877
Qy 121 CAGACATGTTTCATCAATGTTGACTATCTG 149
Db 7878 ATGATGTGTATGTCAACGCTCCCTGACTG 7906

RESULT 14
US-09-270-767-444
; Sequence 444, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
```

; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 444  
; LENGTH: 748  
; TYPE: DNA  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-444

Query Match 6.7%; Score 39.2; DB 4; Length 748;  
Best Local Similarity 48.0%; Pred. No. 0.02;  
Matches 144; Conservative 0; Mismatches 153; Indels 3; Gaps 1;

QY	86	TTGTCTCAGGCGGCTTTGTGATGAAGAAACAGACTCAGACATGTTTCATCAATGTTGACTA	145
Db	8	TTGCCGGAAGCAAGTACATTTCTCAAGACGGACGACATGTTTCATCAATGTGCCCAA	67
QY	146	TCTGACTGAACCTGCTTCTGAAGAAAAACAGAAACACCAACCGTGTTCCTTCTTGAA	205
Db	68	GCTGCTGACCTACCTGGACACGACAGGCAAGCGTACCATATACGTTGTTGGCCAA	127
QY	206	ACTCAATGAGTTTCCCATCAGCAGCCCATTCAGCAAGTGGTTTGTCAAGTAAATCTGAATA	265
Db	128	---GAAGTGAACCGATTTCGCAATAAGAAATCCAAATACTATGTGCGTCGATCAGTT	184
QY	266	TCCGTGGACAGGTACCCACCATTTCTGTCGGCAGCGGTACGTGTTTCTGGCGACGT	325
Db	185	CGCGCGGGAGTATTTCCCGTCTTCCACACTGGACCCGCTACGTACTCACCGGAGACAT	244
QY	326	GGCGAGTCAGGTGATCAATGTCTCAAGAGCGTCCCATACATTAACCTGGAAGACGTGT	385
Db	245	TGTGACGAACCTCTAGTGGCACTCTGAAGACCGTTTATCTGAAGCTGGAGACGTGT	304

RESULT 15

US-09-270-767-15726  
; Sequence 15726, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 15726  
; LENGTH: 748  
; TYPE: DNA  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-15726

Query Match 6.7%; Score 39.2; DB 4; Length 748;  
Best Local Similarity 48.0%; Pred. No. 0.02;  
Matches 144; Conservative 0; Mismatches 153; Indels 3; Gaps 1;

QY	86	TTGTCTCAGGCGGCTTTGTGATGAAGAAACAGACTCAGACATGTTTCATCAATGTTGACTA	145
Db	8	TTGCCGGAAGCAAGTACATTTCTCAAGACGGACGACATGTTTCATCAATGTGCCCAA	67
QY	146	TCTGACTGAACCTGCTTCTGAAGAAAAACAGAAACACCAACCGTGTTCCTTCTTGAA	205
Db	68	GCTGCTGACCTACCTGGACAAAGCAAGGCAAGCGTACCATATACGTTGTTGGCCAA	127
QY	206	ACTCAATGAGTTTCCCATCAGCAGCCCATTCAGCAAGTGGTTTGTCAAGTAAATCTGAATA	265
Db	128	---GAAGTGAACCGATTTCGCAATAAGAAATCCAAATACTATGTGCGTCGATCAGTT	184
QY	266	TCCGTGGACAGGTACCCACCATTTCTGTCGGCAGCGGTACGTGTTTCTGGCGACGT	325
Db	185	CGCGCGGGAGTATTTCCCGTCTTCCACACTGGACCCGCTACGTACTCACCGGAGACAT	244
QY	326	GGCGAGTCAGGTGATCAATGTCTCAAGAGCGTCCCATACATTAACCTGGAAGACGTGT	385

Db 245 TGTGACGAACCTCTACGTGGCATCTCTGAAGACCGTTTATCTGAAGCTGGAGACGTGTT 304

Search completed: September 22, 2005, 17:02:06  
Job time : 123.581 secs

***This Page Blank (uspto)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2005, 14:50:12 ; Search time 1516.01 Seconds  
(without alignments)  
2575.655 Million cell updates/sec

Title: US-10-777-828-8\_COPY\_428\_1011

Perfect score: 584

Sequence: 1 gggacattatccagaaggat.....aagattgcgcgtctgtga 584

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 7400704 seqs, 3343079526 residues

Total number of hits satisfying chosen parameters: 14801408

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*
- 19: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq.\*
- 20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq.\*
- 21: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq.\*
- 22: /cgn2\_6/ptodata/1/pubpna/US10J\_NEW\_PUB.seq.\*
- 23: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq.\*
- 24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 25: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 26: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	584	100.0	1011	19	US-10-777-828-8 Sequence 8, Appli
2	570.4	97.7	681	13	US-10-027-632-21208 Sequence 21208, A
3	570.4	97.7	681	13	US-10-027-632-21209 Sequence 21209, A
4	570.4	97.7	681	13	US-10-027-632-21210 Sequence 21210, A
5	570.4	97.7	681	17	US-10-027-632-21208 Sequence 21208, A
6	570.4	97.7	681	17	US-10-027-632-21209 Sequence 21209, A
7	570.4	97.7	681	17	US-10-027-632-21210 Sequence 21210, A

8	438	75.0	516	13	US-10-027-632-106210	Sequence 106210,
9	438	75.0	516	17	US-10-027-632-106210	Sequence 106210,
10	437.6	74.9	516	13	US-10-027-632-106208	Sequence 106208,
11	437.6	74.9	516	13	US-10-027-632-106208	Sequence 106208,
12	437.6	74.9	516	17	US-10-027-632-106208	Sequence 106208,
13	437.6	74.9	516	17	US-10-027-632-106209	Sequence 106209,
14	119	20.4	1965	21	US-10-764-420-1500	Sequence 1500, Ap
15	114.4	19.6	1266	9	US-09-529-063-26	Sequence 26, Appl
16	114.4	19.6	1266	16	US-10-414-378-26	Sequence 26, Appl
17	114.4	19.6	1269	14	US-10-112-616A-1	Sequence 1, Appl
18	114.4	19.6	1269	21	US-10-764-420-29	Sequence 29, Appl
19	114.4	19.6	2420	9	US-09-529-063-27	Sequence 27, Appl
20	114.4	19.6	2420	16	US-10-414-378-27	Sequence 27, Appl
21	109	18.7	2168	21	US-10-887-553A-603	Sequence 603, App
22	106.2	18.2	1358	20	US-10-370-715B-481	Sequence 481, App
23	106.2	18.2	1358	21	US-10-651-237-77	Sequence 77, Appl
24	106.2	18.2	1358	21	US-10-782-413-77	Sequence 77, Appl
25	106.2	18.2	1773	9	US-09-739-451-9	Sequence 9, Appl
26	106.2	18.2	2095	9	US-09-989-722-208	Sequence 208, App
27	106.2	18.2	2095	9	US-09-989-723-208	Sequence 208, App
28	106.2	18.2	2095	9	US-09-989-279-208	Sequence 208, App
29	106.2	18.2	2095	9	US-09-989-727-208	Sequence 208, App
30	106.2	18.2	2095	9	US-09-989-731-208	Sequence 208, App
31	106.2	18.2	2095	9	US-09-989-732-208	Sequence 208, App
32	106.2	18.2	2095	9	US-09-991-073-208	Sequence 208, App
33	106.2	18.2	2095	9	US-09-990-442-208	Sequence 208, App
34	106.2	18.2	2095	9	US-09-991-163-208	Sequence 208, App
35	106.2	18.2	2095	9	US-09-993-604-208	Sequence 208, App
36	106.2	18.2	2095	9	US-09-990-456-208	Sequence 208, App
37	106.2	18.2	2095	9	US-09-989-721-208	Sequence 208, App
38	106.2	18.2	2095	9	US-09-992-598-208	Sequence 208, App
39	106.2	18.2	2095	9	US-09-989-293A-208	Sequence 208, App
40	106.2	18.2	2095	9	US-09-989-735-208	Sequence 208, App
41	106.2	18.2	2095	9	US-09-990-444-208	Sequence 208, App
42	106.2	18.2	2095	9	US-09-991-181-208	Sequence 208, App
43	106.2	18.2	2095	9	US-09-989-730-208	Sequence 208, App
44	106.2	18.2	2095	9	US-09-990-436-208	Sequence 208, App
45	106.2	18.2	2095	9	US-09-993-687-208	Sequence 208, App

#### ALIGNMENTS

#### RESULT 1

US-10-777-828-8  
; Sequence 8, Application US/10777828  
; Publication No. US20040142425A1  
; GENERAL INFORMATION:  
; APPLICANT: Clausen, Henrik  
; APPLICANT: Amado, Margarita  
; TITLE OF INVENTION: UDP-GALACTOSE: BETA-N-ACETYL-GLUCOSAMINE BETA 1,3  
; TITLE OF INVENTION: GALACTOSYLTRANSFERASES, BETAGAL-T5  
; FILE REFERENCE: 7188-157  
; CURRENT APPLICATION NUMBER: US/10/777,828  
; PRIOR FILING DATE: 2004-02-12  
; PRIOR APPLICATION NUMBER: US/09/831,630  
; PRIOR FILING DATE: 2001-05-10  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 1011  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (79)..(1008)  
US-10-777-828-8

Query Match 100.0%; Score 584; DB 19; Length 1011;  
Best Local Similarity 100.0%; Pred. No. 8.2e-188;  
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGACATTATCCAGAGGATTTCTAGACGCTCTATTACAATCTGACCCCTGAAGACCATGA 60

Db GGGACATATCCAGAGAGATTTCTAGACGCTATTACAACTGACCCCTGAAGACCATGA 487  
Qy TGGGCATAGAAATGGGTCCATCGCTTTTCTCTCAGGCGGCGTTTGTGATGAAACAGACT 120  
Db TGGGCATAGAAATGGGTCCATCGCTTTTCTCTCAGGCGGCGTTTGTGATGAAACAGACT 547  
Qy CAGACATGTTTCAATCAATGTTGACTATCTGACTGAACTGCTTTCTGAAGAAACAGAACAA 180  
Db CAGACATGTTTCAATCAATGTTGACTATCTGACTGAACTGCTTTCTGAAGAAACAGAACAA 607  
Qy CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCCATCAGGCGAGCCATTTCAGCA 240  
Db CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCCATCAGGCGAGCCATTTCAGCA 667  
Qy AGTGGTTTGTCAAGTAAATCTGAATATCCGTGGGACAGGTACCCACCAATTTCTGCTCCGGCA 300  
Db AGTGGTTTGTCAAGTAAATCTGAATATCCGTGGGACAGGTACCCACCAATTTCTGCTCCGGCA 727  
Qy CCGGCTACGTTTCTCGGACAGTGGCGAGTCAAGTGTACAATGTTCTTCCAGAGCGTCC 360  
Db CCGGCTACGTTTCTCGGACAGTGGCGAGTCAAGTGTACAATGTTCTTCCAGAGCGTCC 787  
Qy CATACATTAATACTGGAAGACGCTGTTTGTGGGCTCTGCTCTCGAAAGGCTGAACATCAGAT 420  
Db CATACATTAATACTGGAAGACGCTGTTTGTGGGCTCTGCTCTCGAAAGGCTGAACATCAGAT 847  
Qy TGGAGGAGCTCCACTCCAGCGGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGATATGCC 480  
Db TGGAGGAGCTCCACTCCAGCGGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGATATGCC 907  
Qy TCTTCAGAGAGATCGTGGCCCTGCCACTTCAATCAAGCCCTCGGACTCTCTTGACTACTGGC 540  
Db TCTTCAGAGAGATCGTGGCCCTGCCACTTCAATCAAGCCCTCGGACTCTCTTGACTACTGGC 967  
Qy AGGCTCTAGAGAAATTCGGGGGGGAGAAATTTGCGCCCTGTCTGA 584  
Db AGGCTCTAGAGAAATTCGGGGGGGAGAAATTTGCGCCCTGTCTGA 1011

## RESULT 2

US-10-027-632-21208  
; Sequence 21208, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 21208

LENGTH: 681

TYPE: DNA

ORGANISM: Human

US-10-027-632-21208

Query Match 97.7%; Score 570.4; DB 13; Length 681;  
Best Local Similarity 99.1%; Pred. No. 2.9e-183;  
Matches 571; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GGGACATATCCAGAGAGATTTCTAGACGCTATTACAACTGACCCCTGAAGACCATGA 60  
Db GGGACATATCCAGAGAGATTTCTAGACGCTATTACAACTGACCCCTGAAGACCATGA 135  
Qy 61 TGGGCATAGAAATGGGTCCATCGCTTTTCTCTCAGGCGGCGTTTGTGATGAAACAGACT 120  
Db TGGGCATAGAAATGGGTCCATCGCTTTTCTCTCAGGCGGCGTTTGTGATGAAACAGACT 195  
Qy 121 CAGACATGTTTCAATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAACAGAACAA 180  
Db CAGACATGTTTCAATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAACAGAACAA 255  
Qy 181 CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCCATCAGGCGAGCCATTTCAGCA 240  
Db CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCCATCAGGCGAGCCATTTCAGCA 315  
Qy 241 AGTGGTTTGTCAAGTAAATCTGAATATCCGTGGGACAGGTACCCACCAATTTCTGCTCCGGCA 300  
Db AGTGGTTTGTCAAGTAAATCTGAATATCCGTGGGACAGGTACCCACCAATTTCTGCTCCGGCA 375  
Qy 301 CCGGCTACGTTTCTCGGACAGTGGCGAGTCAAGTGTACAATGTTCTTCCAGAGCGTCC 360  
Db CCGGCTACGTTTCTCGGACAGTGGCGAGTCAAGTGTACAATGTTCTTCCAGAGCGTCC 435  
Qy 361 CATACATTAATACTGGAAGACGCTGTTTGTGGGGCTCTGCCCTCGAAAGGCTGAACATCAGAT 420  
Db CATACATTAATACTGGAAGACGCTGTTTGTGGGGCTCTGCCCTCGAAAGGCTGAACATCAGAT 495  
Qy 421 TGGAGGAGCTCCACTCCAGCGGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGATATGCC 480  
Db TGGAGGAGCTCCACTCCAGCGGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGATATGCC 555  
Qy 481 TCTTCAGAGAGATCGTGGCCCTGCCACTTCAATCAAGCCCTCGGACTCTCTTGACTACTGGC 540  
Db TCTTCAGAGAGATCGTGGCCCTGCCACTTCAATCAAGCCCTCGGACTCTCTTGACTACTGGC 615  
Qy 541 AGGCTCTAGAGAAATTCGGGGGGGAGAAATTTGTCGCG 576  
Db AGGCTCTAGAGAAATTCGGGGGGGAGAAATTTGTCGCG 651

## RESULT 3

US-10-027-632-21209  
; Sequence 21209, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 21209

; LENGTH: 681  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-21209

Query Match 97.7%; Score 570.4; DB 13; Length 681;  
Best Local Similarity 99.1%; Pred. No. 2.9e-183;  
Matches 571; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GGGACATTATCCAGAAAGGATTTCTTAGAGCTCTATTACAATCTGACCCCTGAAGACCATGA 60  
Db GGGACATTATCCAGAAAGGATTTCTTAGAGCTCTATTACAATCTGACCCCTGAAGACCATGA 135  
QY 61 TGGGCATAGAATGGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAACAGACT 120  
Db TGGGCATAGAATGGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAACAGACT 195  
QY 121 CAGACATGTTCAATCAATGTTGACTATCTGACTGAAGTCTTCTGAAGAAACAGAA 180  
Db CAGACATGTTCAATCAATGTTGACTATCTGACTGAAGTCTTCTGAAGAAACAGAA 255  
QY 181 CCAGGTTTTTCACTGGCTTTCTTGAACCTCAATGAGTTTCCCATCAGGCGACCATTCAGCA 240  
Db CCAGGTTTTTCACTGGCTTTCTTGAACCTCAATGAGTTTCCCATCAGGCGACCATTCAGCA 315  
QY 241 AGTGGTTTGTCAAGTAAATCTGAATATCCGTGGGACAGGTACCCACATTCGTCTCCGGCA 300  
Db AGTGGTTTGTCAAGTAAATCTGAATATCCGTGGGACAGGTACCCACATTCGTCTCCGGCA 375  
QY 301 CCGGCTACGTGTTTCTGGCAGCTGGCGAGTCAGGTGTACAATCTCTCCAAAGCGCTCC 360  
Db CCGGCTACGTGTTTCTGGCAGCTGGCGAGTCAGGTGTACAATCTCTCCAAAGCGCTCC 435  
QY 361 CATACATTAAACTGGAAGACGCTGTTTGTGGGCTCTGCTCGAAAGGCTGAACATCAGAT 420  
Db CATACATTAAACTGGAAGACGCTGTTTGTGGGCTCTGCTCGAAAGGCTGAACATCAGAT 495  
QY 421 TGGAGGAGCTCCACTCCAGCGACCTTTTTCAGGGGGCTTACGCTTCTCCGATGCC 480  
Db TGGAGGAGCTCCACTCCAGCGACCTTTTTCAGGGGGCTTACGCTTCTCCGATGCC 555  
QY 481 TCTTCAGAGGATCTGGCTGCCACTTCATCAAGCCTCGGACTCTCTGGACTACTGGC 540  
Db TCTTCAGAGGATCTGGCTGCCACTTCATCAAGCCTCGGACTCTCTGGACTACTGGC 615  
QY 541 AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCGCG 576  
Db AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCGCG 651

## RESULT 4

US-10-027-632-21210  
; Sequence 21210, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827,129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 21210  
; LENGTH: 681  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-21210

Query Match 97.7%; Score 570.4; DB 13; Length 681;  
Best Local Similarity 99.1%; Pred. No. 2.9e-183;  
Matches 571; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GGGACATTATCCAGAAAGGATTTCTTAGAGCTCTATTACAATCTGACCCCTGAAGACCATGA 60  
Db GGGACATTATCCAGAAAGGATTTCTTAGAGCTCTATTACAATCTGACCCCTGAAGACCATGA 135  
QY 61 TGGGCATAGAATGGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAACAGACT 120  
Db TGGGCATAGAATGGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAACAGACT 195  
QY 121 CAGACATGTTCAATCAATGTTGACTATCTGACTGAAGTCTTCTGAAGAAACAGAA 180  
Db CAGACATGTTCAATCAATGTTGACTATCTGACTGAAGTCTTCTGAAGAAACAGAA 255  
QY 181 CCAGGTTTTTCACTGGCTTTCTTGAACCTCAATGAGTTTCCCATCAGGCGACCATTCAGCA 240  
Db CCAGGTTTTTCACTGGCTTTCTTGAACCTCAATGAGTTTCCCATCAGGCGACCATTCAGCA 315  
QY 241 AGTGGTTTGTCAAGTAAATCTGAATATCCGTGGGACAGGTACCCACATTCGTCTCCGGCA 300  
Db AGTGGTTTGTCAAGTAAATCTGAATATCCGTGGGACAGGTACCCACATTCGTCTCCGGCA 375  
QY 301 CCGGCTACGTGTTTCTGGCAGCTGGCGAGTCAGGTGTACAATCTCTCCAAAGCGCTCC 360  
Db CCGGCTACGTGTTTCTGGCAGCTGGCGAGTCAGGTGTACAATCTCTCCAAAGCGCTCC 435  
QY 361 CATACATTAAACTGGAAGACGCTGTTTGTGGGCTCTGCTCGAAAGGCTGAACATCAGAT 420  
Db CATACATTAAACTGGAAGACGCTGTTTGTGGGCTCTGCTCGAAAGGCTGAACATCAGAT 495  
QY 421 TGGAGGAGCTCCACTCCAGCGACCTTTTTCAGGGGGCTTACGCTTCTCCGATGCC 480  
Db TGGAGGAGCTCCACTCCAGCGACCTTTTTCAGGGGGCTTACGCTTCTCCGATGCC 555  
QY 481 TCTTCAGAGGATCTGGCTGCCACTTCATCAAGCCTCGGACTCTCTGGACTACTGGC 540  
Db TCTTCAGAGGATCTGGCTGCCACTTCATCAAGCCTCGGACTCTCTGGACTACTGGC 615  
QY 541 AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCGCG 576  
Db AGGCTCTAGAGAAATCCCGGGGGGAAGATTGTCGCG 651

## RESULT 5

US-10-027-632-21208  
; Sequence 21208, Application US/10027632  
; Publication No. US20030204075N9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827,129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218

```

; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: US 60/167,363
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 21208
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-21208

Query Match      97.7%   Score 570.4;   DB 17;   Length 681;
Best Local Similarity 99.1%   Pred. No. 2.9e-183;
Matches 571;   Conservative 2;   Mismatches 3;   Indels 0;   Gaps 0;

Qy 1 GGGACATTATCCAGAAGGATTTCCTAGACGCTATTACAATCTGACCTGAGACCATGA 60
Db 76 GGGACATTATCCAGAAGGATTTCCTAGACGCTATTACAATCTGACCTGAGACCATGA 135

Qy 61 TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGCGGCGTTTGTGATGAAACAGACT 120
Db 136 TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGCGGCGTTTGTGATGAAACAGACT 195

Qy 121 CAGACATGTTCAATCAATGTTGACTATCTGACTGAACTGCTCTGAAAGAAAACAGAACAA 180
Db 196 CAGACATGTTCAATCAATGTTGACTATCTGACTGAACTGCTCTGAAAGAAAACAGAACAA 255

Qy 181 CCAGGTTTTTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGCGGCGTTTGTGATGAAACAGACT 240
Db 256 CCAGGTTTTTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGCGGCGTTTGTGATGAAACAGACT 315

Qy 241 AGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCA 300
Db 316 AGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCA 375

Qy 301 CCGGCTACGTTGTTCTGGCGACGTGGCGAGTCAAGTGTACAAATGTTCTGCTCCGGCA 360
Db 376 CCGGCTACGTTGTTCTGGCGACGTGGCGAGTCAAGTGTACAAATGTTCTGCTCCGGCA 435

Qy 361 CATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT 420
Db 436 CATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT 495

Qy 421 TCTTCAGAGAGATCGTGCGCTGCCACTTTCATCAAGCCCTCGGACTCTCTTGGACTACTGGC 480
Db 496 TCTTCAGAGAGATCGTGCGCTGCCACTTTCATCAAGCCCTCGGACTCTCTTGGACTACTGGC 555

Qy 541 AGGCTCTAGAGAAATTCGCGGGGGAAGATTGTCGCG 576
Db 616 AGGCTCTAGAGAAATTCGCGGGGGAAGATTGTCGCGCG 651

```

```

RESULT 6
US-10-027-632-21209
; Sequence 21209, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12

```

```

; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 21209
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-21209

Query Match      97.7%   Score 570.4;   DB 17;   Length 681;
Best Local Similarity 99.1%   Pred. No. 2.9e-183;
Matches 571;   Conservative 2;   Mismatches 3;   Indels 0;   Gaps 0;

Qy 1 GGGACATTATCCAGAAGGATTTCCTAGACGCTATTACAATCTGACCTGAGACCATGA 60
Db 76 GGGACATTATCCAGAAGGATTTCCTAGACGCTATTACAATCTGACCTGAGACCATGA 135

Qy 61 TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGCGGCGTTTGTGATGAAACAGACT 120
Db 136 TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGGCGGCGTTTGTGATGAAACAGACT 195

Qy 121 CAGACATGTTCAATCAATGTTGACTATCTGACTGAACTGCTCTGAAAGAAAACAGAACAA 180
Db 196 CAGACATGTTCAATCAATGTTGACTATCTGACTGAACTGCTCTGAAAGAAAACAGAACAA 255

Qy 181 CCAGGTTTTTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGCGGCGTTTGTGATGAAACAGACT 240
Db 256 CCAGGTTTTTCACTGGCTTCTTGAACCTCAATGAGTTTCCCATCAGGCGGCGTTTGTGATGAAACAGACT 315

Qy 241 AGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCA 300
Db 316 AGTGGTTTGTGAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCA 375

Qy 301 CCGGCTACGTTGTTCTGGCGACGTGGCGAGTCAAGTGTACAAATGTTCTGCTCCGGCA 360
Db 376 CCGGCTACGTTGTTCTGGCGACGTGGCGAGTCAAGTGTACAAATGTTCTGCTCCGGCA 435

Qy 361 CATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT 420
Db 436 CATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT 495

Qy 421 TGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTCTCCGATATGCC 480
Db 496 TGGAGGAGCTCCACTCCAGCGGACCTTTTTCAGGGGGCTTACGCTTCTCCGATATGCC 555

Qy 481 TCTTCAGAGAGATCGTGCGCTGCCACTTTCATCAAGCCCTCGGACTCTCTTGGACTACTGGC 540
Db 556 TCTTCAGAGAGATCGTGCGCTGCCACTTTCATCAAGCCCTCGGACTCTCTTGGACTACTGGC 615

Qy 541 AGGCTCTAGAGAAATTCGCGGGGGAAGATTGTCGCG 576
Db 616 AGGCTCTAGAGAAATTCGCGGGGGAAGATTGTCGCGCG 651

```

```

RESULT 7
US-10-027-632-21210
; Sequence 21210, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome

```

RESULT 8  
US-10-027-632-106210  
: Sequence 106210, Application US/10027632

RESULT 9  
US-10-027-632-106210  
; Sequence 106210, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.

```

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 106210
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-106210

```

```

Query Match      75.0%; Score 438; DB 17; Length 516;
Best Local Similarity 100.0%; Pred. No. 3.5e-138;
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGACATTATCCAGAAGGATTTCTAGACGCTATTACAACTGACCTGAGACCATGA 60
Db 79 GGGACATTATCCAGAAGGATTTCTAGACGCTATTACAACTGACCTGAGACCATGA 138

Qy 61 TGGGCATAGAAATGGGTCCATCGCTTTTCTCCTCAGCGCGCGTTTGTGATGAAACAGACT 120
Db 139 TGGGCATAGAAATGGGTCCATCGCTTTTCTCCTCAGCGCGCGTTTGTGATGAAACAGACT 198

Qy 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTCAATGAGTTTCCATCAGGAGCAAA 180
Db 199 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTCAATGAGTTTCCATCAGGAGCAAA 258

Qy 181 CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCATCAGGAGCAAA 240
Db 259 CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCATCAGGAGCAAA 318

Qy 241 AGTGGTTTGTTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCA 300
Db 319 AGTGGTTTGTTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCA 378

Qy 301 CCGGCTACGTGTTTCTGGGACAGTGGCGAGTCAATGTTTCCATCAGGAGCAAA 360
Db 379 CCGGCTACGTGTTTCTGGGACAGTGGCGAGTCAATGTTTCCATCAGGAGCAAA 438

Qy 361 CATACATTAAACTGGAAGACGTTGTTGTGGGGCTCTGCCCTCGAAAGGCTGAACATCAGAT 420
Db 439 CATACATTAAACTGGAAGACGTTGTTGTGGGGCTCTGCCCTCGAAAGGCTGAACATCAGAT 498

Qy 421 TGGAGGAGCTCCACTCCC 438
Db 499 TGGAGGAGCTCCACTCCC 516

```

```

RESULT 10
US-10-027-632-106210
; Sequence 106208, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006

```

```

; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 106208
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-106208

```

```

Query Match      74.9%; Score 437.6; DB 13; Length 516;
Best Local Similarity 99.8%; Pred. No. 4.8e-138;
Matches 437; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGACATTATCCAGAAGGATTTCTAGACGCTATTACAACTGACCTGAGACCATGA 60
Db 79 GGGACATTATCCAGAAGGATTTCTAGACGCTATTACAACTGACCTGAGACCATGA 138

Qy 61 TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGCGCGCGTTTGTGATGAAACAGACT 120
Db 139 TGGGCATAGAAATGGGTCCATCGCTTTTGTCTCAGCGCGCGTTTGTGATGAAACAGACT 198

Qy 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTCAATGAGTTTCCATCAGGAGCAAA 180
Db 199 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTCAATGAGTTTCCATCAGGAGCAAA 258

Qy 181 CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCATCAGGAGCAAA 240
Db 259 CCAGGTTTTTCACTGGCTTCTTGAACACTCAATGAGTTTCCATCAGGAGCAAA 318

Qy 241 AGTGGTTTGTTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCA 300
Db 319 AGTGGTTTGTTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCA 378

Qy 301 CCGGCTACGTGTTTCTGGGACAGTGGCGAGTCAATGTTTCCATCAGGAGCAAA 360
Db 379 CCGGCTACGTGTTTCTGGGACAGTGGCGAGTCAATGTTTCCATCAGGAGCAAA 438

Qy 361 CATACATTAAACTGGAAGACGTTGTTGTGGGGCTCTGCCCTCGAAAGGCTGAACATCAGAT 420
Db 439 CATACATTAAACTGGAAGACGTTGTTGTGGGGCTCTGCCCTCGAAAGGCTGAACATCAGAT 498

Qy 421 TGGAGGAGCTCCACTCCC 438
Db 499 TGGAGGAGCTCCACTCCC 516

```

```

RESULT 11
US-10-027-632-106209
; Sequence 106209, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006

```

```
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 106209
/ LENGTH: 516
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-106209
```

```
Query Match 74.9%; Score 437.6; DB 13; Length 516;
Best Local Similarity 99.8%; Pred. No. 4.8e-138;
Matches 437; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGACATTATCCAGAAGGATTTCTAGAGCTCTATTACAATCTGACCCCTGAAGACCATGA 60
DB 79 GGGACATTATCCAGAAGGATTTCTAGAGCTCTATTACAATCTGACCCCTGAAGACCATGA 138

QY 61 TGGGCATAGATGGGTCCATCGCTTTTGTCTCAGCGCGGTTTGTGATGAAAAACAGACT 120
DB 139 TGGGCATAGATGGGTCCATCGCTTTTGTCTCAGCGCGGTTTGTGATGAAAAACAGACT 198

QY 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTCTCTGAGAAACAGAACAA 180
DB 199 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTCTCTGAGAAACAGAACAA 258

QY 181 CCAGGTTTTTCACTGGCTTTTGAACAATCAATGAGTTTCCCATCAGGAGCCATTTCAGCA 240
DB 259 CCAGGTTTTTCACTGGCTTTTGAACAATCAATGAGTTTCCCATCAGGAGCCATTTCAGCA 318

QY 241 AGTGGTTTGTGAGTAAATCTGAATATCGTGGGACAGGTACCCACATTTCTGCTCCGGCA 300
DB 319 AGTGGTTTGTGAGTAAATCTGAATATCGTGGGACAGGTACCCACATTTCTGCTCCGGCA 378

QY 301 CCGGCTACGTGTTTCTGCGCAGCGTGGGAGTCAGGTGTACAAATGTTCTCCAGAGCGTCC 360
DB 379 CCGGCTACGTGTTTCTGCGCAGCGTGGGAGTCAGGTGTACAAATGTTCTCCAGAGCGTCC 438

QY 361 CATACATTAAACTGGAAGACGTTTGTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT 420
DB 439 CATACATTAAACTGGAAGACGTTTGTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT 498

QY 421 TGGAGGAGCTCCACTCCC 438
DB 499 TGGAGGAGCTCCACTCCC 516
```

```
RESULT 12
US-10-027-632-106209
/ Sequence 106208, Application US/10027632
/ Publication No. US20030204075A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ PRIOR FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
```

```
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 106208
/ LENGTH: 516
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-106208

Query Match 74.9%; Score 437.6; DB 17; Length 516;
Best Local Similarity 99.8%; Pred. No. 4.8e-138;
Matches 437; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGACATTATCCAGAAGGATTTCTAGAGCTCTATTACAATCTGACCCCTGAAGACCATGA 60
DB 79 GGGACATTATCCAGAAGGATTTCTAGAGCTCTATTACAATCTGACCCCTGAAGACCATGA 138

QY 61 TGGGCATAGATGGGTCCATCGCTTTTGTCTCAGCGCGGTTTGTGATGAAAAACAGACT 120
DB 139 TGGGCATAGATGGGTCCATCGCTTTTGTCTCAGCGCGGTTTGTGATGAAAAACAGACT 198

QY 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTCTCTGAGAAACAGAACAA 180
DB 199 CAGACATGTTTCATCAATGTTGACTATCTGACTGAACTCTCTGAGAAACAGAACAA 258

QY 181 CCAGGTTTTTCACTGGCTTTTGAACAATCAATGAGTTTCCCATCAGGAGCCATTTCAGCA 240
DB 259 CCAGGTTTTTCACTGGCTTTTGAACAATCAATGAGTTTCCCATCAGGAGCCATTTCAGCA 318

QY 241 AGTGGTTTGTGAGTAAATCTGAATATCGTGGGACAGGTACCCACATTTCTGCTCCGGCA 300
DB 319 AGTGGTTTGTGAGTAAATCTGAATATCGTGGGACAGGTACCCACATTTCTGCTCCGGCA 378

QY 301 CCGGCTACGTGTTTCTGCGCAGCGTGGGAGTCAGGTGTACAAATGTTCTCCAGAGCGTCC 360
DB 379 CCGGCTACGTGTTTCTGCGCAGCGTGGGAGTCAGGTGTACAAATGTTCTCCAGAGCGTCC 438

QY 361 CATACATTAAACTGGAAGACGTTTGTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT 420
DB 439 CATACATTAAACTGGAAGACGTTTGTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT 498

QY 421 TGGAGGAGCTCCACTCCC 438
DB 499 TGGAGGAGCTCCACTCCC 516
```

```
RESULT 13
US-10-027-632-106209
/ Sequence 106209, Application US/10027632
/ Publication No. US20030204075A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ PRIOR FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
```

; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 106209  
; LENGTH: 516  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-106209

Query Match 74.9%; Score 437.6; DB 17; Length 516;  
Best Local Similarity 99.8%; Pred. No. 4.8e-138;  
Matches 437; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGACATTATCCAGAAGGATTTCTAGACGCTATTACAATCTGACCTGAGACCATGA 60  
Db 75 GGGACATTATCCAGAAGGATTTCTAGACGCTATTACAATCTGACCTGAGACCATGA 138  
Qy 61 TGGGCATAGAATGGTCCATCGCTTTTGTCTCAGGCGGCGTTTGTGATGAAAAACAGACT 120  
Db 139 TGGGCATAGAATGGTCCATCGCTTTTGTCTCAGGCGGCGTTTGTGATGAAAAACAGACT 198  
Qy 121 CAGACATTTTCATCAATGTTGACTATCTGACTGATGAACTGCTTCTGAAGAAAAACAGAA 180  
Db 199 CAGACATTTTCATCAATGTTGACTATCTGACTGATGAACTGCTTCTGAAGAAAAACAGAA 258  
Qy 181 CCAGGTTTTTCACTGGCTTCTTGAACCTCAATGATGTTTCCCATCAGGCGAGCATTCAGCA 240  
Db 259 CCAGGTTTTTCACTGGCTTCTTGAACCTCAATGATGTTTCCCATCAGGCGAGCATTCAGCA 318  
Qy 241 AGTGGTTTGTGAGTAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCA 300  
Db 378 AGTGGTTTGTGAGTAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCA 378  
Qy 301 CCGGCTACGTGTTTCTGGGACGTGGCGAGTCAAGTGTACAATGTCTTCCAAGAGCGTCC 360  
Db 379 CCGGCTACGTGTTTCTGGGACGTGGCGAGTCAAGTGTACAATGTCTTCCAAGAGCGTCC 438  
Qy 361 CATACATTAACTGGAAGACGTGTTTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT 420  
Db 439 CATACATTAACTGGAAGACGTGTTTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT 498  
Qy 421 TGGAGGAGCTCCACTCCC 438  
Db 439 TGGAGGAGCTCCACTCCC 516

RESULT 14  
US-10-764-420-1500  
; Sequence 150, Application US/10764420  
; Publication No. US2005008472A1  
; GENERAL INFORMATION:  
; APPLICANT: Lum, Pek Yee  
; APPLICANT: Tan, Yejun  
; APPLICANT: Dai, Hongyue  
; TITLE OF INVENTION: Methods For Determining Whether An Agent Possesses A Defined Biological Activity  
; FILE REFERENCE: ROSA122057  
; CURRENT APPLICATION NUMBER: US/10/764,420  
; PRIOR FILING DATE: 2004-01-23  
; PRIOR APPLICATION NUMBER: US 60/442,797  
; PRIOR FILING DATE: 2003-01-24  
; PRIOR APPLICATION NUMBER: US 60/474,413  
; PRIOR FILING DATE: 2003-05-30  
; NUMBER OF SEQ ID NOS: 3683  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1500  
; LENGTH: 1965

; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-764-420-1500

Query Match 20.4%; Score 119; DB 21; Length 1965;  
Best Local Similarity 51.2%; Pred. No. 4.4e-29;  
Matches 278; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

Qy 1 GGGACATTATCCAGAAGGATTTCTAGACGCTATTACAATCTGACCTGAGACCATGA 60  
Db 646 GTGATATTATACGGCAAGACTTTCTAGACACATATAATAACTTTGACCTTGAACACCATTA 705  
Qy 61 TGGGCATAGAATGGTCCATCGCTTTTGTCTCAGGCGGCGTTTGTGATGAAAAACAGACT 120  
Db 706 TGGCTTTTCAGGTGGTAAATGGAGTTTGGCCCAATGCCAAGTATATTATGAAACACACA 765  
Qy 121 CAGACATGTTTCATCAATGTTGACTATCTGACTGAATGCTTCTGAAGAAAAACAGAA 180  
Db 766 CTGATGTTTTCATCAATGTTGCAATTTAGTCTATTATCTTTTAAACCTAAACCACTCAG 825  
Qy 181 CCAGGTTTTTCACTGGCTTCTTGAACCTCAATGATGTTTCCCATCAGGCGAGCATTCAGCA 240  
Db 826 AGAAGTTTTTTCACGGGCTATCTCTAATGATAACTATCTATAGAGGATTTTTCATA 885  
Qy 241 AGTGGTTTGTGAGTAATCTGAATATCCGTGGGACAGGTACCCACCATTTCTGCTCCGGCA 300  
Db 886 ARAACACATTTTCATCAAGAGTACCCCTTCAAGGTGTTCCCTCTACTGACAGCGGC 945  
Qy 301 CCGGCTACGTGTTTCTGGGACGTGGCGAGTCAAGTGTACAATGTCTTCCAAGAGCGTCC 360  
Db 946 TGGGCTACATTATGTCGCGGCGACTGCTGTCGCCAGGCTTACGAGATGATGATGATCAGCTGA 1005  
Qy 361 CATACATTAACTGGAAGACGTGTTTGGGGCTCTGCTCGAAAGGCTGAACATCAGAT 420  
Db 1006 AGCCCATCAAGTTTGAAGACGTTTATGTTGGCATCTGTTTGAATTTGTTAAAGTGGACA 1065  
Qy 421 TGGAGGAGCTCCACTCCCGACGACCTTTTTCAGGGGCTTACGCTTCTCCGATATGCC 480  
Db 1066 TTCTATTTCCAGAGACACAAACCTTTTCTTCTGTACAGATCCACTTGGATGATGTC 1125  
Qy 481 TCTTCAGAGAGATGTTGGGCTGGCCTGCACTTCATCAAGCTTCGGAGCTCTCTTGGAGCTACTGGC 540  
Db 1126 AGCTCAGACGCGTATGTCAGCCCATGGCTTTTCTTCCAAGGAGATCATCAGATTTCTGGC 1185  
Qy 541 AGG 543  
Db 1186 AGG 1188

RESULT 15  
US-09-529-063-26  
; Sequence 26, Application US/09529063  
; Patent No. US20020102542A1  
; GENERAL INFORMATION:  
; APPLICANT: FUKUSHIMA, DAIKICHI  
; APPLICANT: SHIBAYAMA, SHIRO  
; APPLICANT: TADA, HIDEAKI  
; TITLE OF INVENTION: POLYPEPTIDE, cDNA ENCODING THE POLYPEPTIDE, AND USE OF  
; FILE REFERENCE: Q58769  
; CURRENT APPLICATION NUMBER: US/09/529,063  
; CURRENT FILING DATE: 2000-04-07  
; PRIOR FILING DATE: 1998-10-06  
; PRIOR APPLICATION NUMBER: JP 9-274674  
; PRIOR FILING DATE: 1997-10-07  
; NUMBER OF SEQ ID NOS: 117  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 26  
; LENGTH: 1266  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-529-063-26



Search completed: September 22, 2005, 22:20:25  
Job time : 1517.01 secs

**This Page Blank (uspto)**